

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Barenkamp, Stephen J
- (ii) TITLE OF INVENTION: High Molecular Weight Surface Proteins of Non-Typeable Haemophilus
- (iii) NUMBER OF SEQUENCES: 11
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Shoemaker and Mattare, Ltd.
 - (B) STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza Bldg. 1
 - (C) CITY: Arlington
 - (D) STATE: Virginia
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 22202-0286
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/617,697
 - (B) FILING DATE: 01-APR-1996
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/302,832
 - (B) FILING DATE: 05-OCT-1994
- (viii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US PCT/US93/02166
 - (B) FILING DATE: 16-MAR-1993
- (ix) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Berkstresser, Jerry W
 - (B) REGISTRATION NUMBER: 22,651
 - (C) REFERENCE/DOCKET NUMBER: 1038-557
- (x) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (703) 415-0810
 - (B) TELEFAX: (703) 415-0813

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5116 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTT GCAGTCTATA TGCAAATATT TAAAAAAATA GTATAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTCA TCTTCATCT TTCATCTTCA ATCTTCATC	180

TTTCATCTT	CATCTTCAT	CTTCATCTT	TCATCTTC	TCTTCATCT	TTCATCTTC	240
ACATGCCCTG	ATGAACCGAG	GGAAGGGAGG	GAGGGGCAAG	AATGAAGAGG	GAGCTGAACG	300
AACGCAAATG	ATAAAAGTAAT	TTAATTGTT	AACTAACCTT	AGGAGAAAAT	ATGAACAAGC	360
TATATCGTCT	CAAATTCA	AAACGCCTGA	ATGCTTGTT	TGCTGTGTCT	GAATTGGCAC	420
GGGGTTGTGA	CCATTCCACA	GAAAAAGGCA	GCGAAAAAACC	TGCTCGCATG	AAAGTGCCTC	480
ACTTAGCGTT	AAAGCCACTT	TCCGCTATGT	TACTATCTT	AGGTGTAACA	TCTATTCCAC	540
AATCTGTTT	AGCAAGCGGC	TTACAAGGAA	TGGATGTTAGT	ACACGGCACA	GCCACTATGC	600
AAGTAGATGG	TAATAAAACC	ATTATCCGCA	ACAGTGTG	CGATATCATT	AATTGGAAAC	660
AATTTAACAT	CGACCAAAT	GAAATGGTGC	AGTTTTTACA	AGAAAACAAC	AACTCCGCCG	720
TATTCAACCG	TGTTACATCT	AACCAAATCT	CCCAATTAAA	AGGGATTTA	GATTCTAACG	780
GACAAGTCTT	TTTAATCAAC	CCAAATGGTA	TCACAATAGG	TAAAGACGCA	ATTATTAACA	840
CTAATGGCTT	TACGGCTTCT	ACGCTAGACA	TTCTAACGA	AAACATCAAG	GCCGTAATT	900
TCACCTTCGA	GCAAACCAAA	GATAAAGCGC	TCGCTGAAAT	TGTGAATCAC	GGTTTAATT	960
CTGTCGGTAA	AGACGGCAGT	GTAAATCTT	TTGGTGGCAA	AGTAAAAAC	GAGGGTGTGA	1020
TTAGCGTAA	TGGTGGCAGC	ATTTCTTAC	TCGCAGGGCA	AAAAATCACC	ATCAGCGATA	1080
TAATAAACCC	AACCATTACT	TACAGCATTG	CCGCGCCTGA	AAATGAAGCG	GTCAATCTGG	1140
GCGATATTTT	TGCCAAAGGC	GGTAACATTA	ATGTCGTGC	TGCCACTATT	CGAAACCAAG	1200
GTAAACTTTC	TGCTGATTCT	GTAAGCAAAG	ATAAAAGCGG	CAATATTGTT	CTTCCGCCA	1260
AAGAGGGTGA	AGCGGAAATT	GGCGGTGTA	TTTCCGCTCA	AAATCAGCAA	GCTAAAGCG	1320
GCAAGCTGAT	GATTACAGGC	GATAAAGTC	CATTAAAAC	AGGTGCAGTT	ATCGACCTTT	1380
CAGGTAAAGA	AGGGGGAGAA	ACTTACCTTG	GGGGTGACGA	GCGCGCGAA	GGTAAAAGG	1440
GCATTCAATT	AGCAAAGAAA	ACCTCTT	AAAAAGGCTC	AACCATCAAT	GTATCAGGCA	1500
AAGAAAAAGG	CGGACGCGCT	ATTGTGTGGG	GGCATATTGC	GTAAATTGAC	GGCAATATTA	1560
ACGCTCAAGG	TAGTGGTGAT	ATCGCTAAA	CCGGTGGTTT	TGTGGAGACG	TCGGGGCATG	1620
ATTTATTCA	CAAAGACAAT	GCAATTGTT	ACGCCAAAGA	GTGGTTGTTA	GACCCGGATA	1680
ATGTATCTAT	TAATGCAGAA	ACAGCAGGAC	GCAGCAATAC	TTCAGAAGAC	GATGAATACA	1740
CGGGATCCGG	GAATAGTGCC	AGCACCCAA	AACGAAACAA	AGAAAAGACA	ACATTAACAA	1800
ACACAAC	TGAGAGTATA	CTAAAAAAAG	GTACCTTGT	TAACATCACT	GCTAATCAAC	1860
GCATCTATGT	CAATAGCTCC	ATTAATTAT	CCAATGGCAG	CTTAACTCTT	TGGAGTGAGG	1920
GTCGGAGCGG	TGGCGCGTT	GAGATTAACA	ACGATATTAC	CACCGGTGAT	GATACCAGAG	1980
GTGCAAACCTT	AACAATTAC	TCAGGGGCT	GGGTTGATGT	TCATAAAAAT	ATCTCACTCG	2040
GGGCGCAAGG	TAACATAAAC	ATTACAGCTA	AACAAGATAT	CGCCTTGAG	AAAGGAAGCA	2100
ACCAAGTCAT	TACAGGTCAA	GGGACTATTA	CCTCAGGCAA	TCAAAAAGGT	TTTAGATT	2160
ATAATGTCTC	TCTAAACGGC	ACTGGCAGCG	GACTGCAATT	CACCACTAAA	AGAACCAATA	2220

AATACGCTAT	CACAAATAAA	TTTGAAGGGA	CTTTAAATAT	TTCAGGGAAA	GTGAACATCT	2280
CAATGGTTTT	ACCTAAAAAT	GAAAGTGGAT	ATGATAAATT	CAAAGGACGC	ACTTACTGGA	2340
ATTTAACCTC	CTTAAATGTT	TCCGAGAGTG	GCGAGTTAA	CCTCACTATT	GACTCCAGAG	2400
GAAGCGATAG	TGCAGGCACA	CTTACCCAGC	CTTATAATT	AAACGGTATA	TCATTCAACA	2460
AAGACACTAC	CTTTAATGTT	GAACGAAATG	CAAGAGTCAA	CTTGACATC	AAGGCACCAA	2520
TAGGGATAAA	TAAGTATTCT	AGTTGAATT	ACGCATCATT	TAATGGAAAC	ATTCAGTTT	2580
CGGGAGGGGG	GAGTGTGAT	TTCACACTTC	TCGCCTCATC	CTCTAACGTC	CAAACCCCCG	2640
GTGTAGTTAT	AAATTCTAAA	TACTTTAATG	TTTCAACAGG	GTCAAGTTA	AGATTTAAAA	2700
CTTCAGGCTC	AAACAAAAC	GGCTCTCAA	TAGAGAAAGA	TTTAACTTTA	AATGCCACCG	2760
GAGGCAACAT	AAACACTTTG	CAAGTTGAAG	GCACCGATGG	AATGATTGGT	AAAGGCATTG	2820
TAGCCAAAAA	AAACATAACC	TTTGAAGGAG	GTAACATCAC	CTTGGCTCC	AGGAAAGCCG	2880
TAACAGAAAT	CGAAGGCAAT	GTTACTATCA	ATAACAACGC	TAACGTCACT	CTTATCGGTT	2940
CGGATTTGA	CAACCATCAA	AAACCTTTAA	CTATTAAAAA	AGATGTCATC	ATTAATAGCG	3000
GCAACCTTAC	CGCTGGAGGC	AATATTGTCA	ATATAGCCGG	AAATCTTACC	GTTGAAAGTA	3060
ACGCTAATTT	CAAAGCTATC	ACAAATTCA	CTTTAATGT	AGGCGGCTTG	TTTGACAACA	3120
AAGGCAATT	AAATATTTC	ATTGCCAAAG	GAGGGGCTCG	CTTTAAAGAC	ATTGATAATT	3180
CCAAGAATT	AAGCATTCAACC	ACCAACTCCA	GCTCCACTTA	CCGCACTATT	ATAAGCGGCA	3240
ATATAACAA	AAAAAACGGT	GATTAAATA	TTACGAACGA	AGGTAGTGAT	ACTGAAATGC	3300
AAATTGGCGG	CGATGTCTCG	AAAAAGAAG	GTAATCTCAC	GATTCTTCT	GACAAATCA	3360
ATATTACAA	ACAGATAACA	ATCAAGGCAG	GTGTTGATGG	GGAGAATTCC	GATTCAAGACG	3420
CGACAAACAA	TGCCAATCTA	ACCATTAAAA	CCAAAGAATT	GAAATTAAACG	CAAGACCTAA	3480
ATATTCAGG	TTTCAATAAA	GCAGAGATTA	CAGCTAAAGA	TGGTAGTGAT	TTAACTATTG	3540
GTAACACCAA	TAGTGCTGAT	GGTACTAATG	CCAAAAAAAGT	AACCTTTAAC	CAGGTTAAAG	3600
ATTCAAAAAT	CTCTGCTGAC	GGTCACAAGG	TGACACTACA	CAGCAAAGTG	GAACATCCG	3660
GTAGTAATAA	CAACACTGAA	GATAGCAGTG	ACAATAATGC	CGGCTTAAC	ATCGATGCAA	3720
AAAATGTAAC	AGTAAACAAAC	AATATTACTT	CTCACAAAGC	AGTGAGGCATC	TCTGCGACAA	3780
GTGGAGAAAT	TACCACTAAA	ACAGGTACAA	CCATTAACGC	AACCACTGGT	AACGTGGAGA	3840
TAACCGCTCA	AACAGGTAGT	ATCCTAGGTG	GAATTGAGTC	CAGCTCTGGC	TCTGTAACAC	3900
TTACTGCAAC	CGAGGGCGCT	CTTGCTGTAA	GCAATATTTC	GGGCAACACC	GTTACTGTTA	3960
CTGCAAATAG	CGGTGCATTA	ACCACTTGG	CAGGCTCTAC	AATTAAAGGA	ACCGAGAGTG	4020
TAACCACTTC	AAGTCAATCA	GGCGATATCG	GCGGTACGAT	TTCTGGTGGC	ACAGTAGAGG	4080
TTAAAGCAAC	CGAAAGTTA	ACCACTCAAT	CCAATTCAA	AATTAAAGCA	ACAACAGGCG	4140
AGGCTAACGT	AACAAGTGCA	ACAGGTACAA	TTGGTGGTAC	GATTCCGGT	AATACGGTAA	4200
ATGTTACGGC	AAACGCTGGC	GATTTAACAG	TTGGGAATGG	CGCAGAAATT	AATGCGACAG	4260

AAGGAGCTGC AACCTTAAC	ACATCATCGG GCAAATTAAC	TACCGAAGCT AGTTCACACA	4320
TTACTTCAGC CAAGGGTCAG	GTAAATCTT CAGCTCAGGA	TGGTAGCGTT GCAGGAAGTA	4380
TTAATGCCGC CAATGTGACA	CTAAATACTA CAGGCACCTT	AACTACCGTG AAGGGTTCAA	4440
ACATTAATGC AACCAAGCGGT	ACCTTGGTTA TTAACGCAA	AGACGCTGAG CAAATGGCG	4500
CAGCATTGGG TAACCACACA	GTGGTAAATG CAACCAACGC	AAATGGCTCC GGCAGCGTAA	4560
TCGCGACAAC CTCAGCAGA	GTGAACATCA CTGGGGATT	AATCACAATA AATGGATTAA	4620
ATATCATTTC AAAAAACGGT	ATAAACACCG TACTGTTAA	AGGCAGTTAA ATTGATGTGA	4680
AATACATTCA ACCGGGTATA	GCAAGCGTAG ATGAAGTAAT	TGAAGCGAAA CGCATTCTTG	4740
AGAAGGTAAA AGATTTATCT	GATGAAGAAA GAGAAGCGTT	AGCTAAACTT GGAGTAAGTG	4800
CTGTACGTTT TATTGAGCCA	AATAATACAA TTACAGTCGA	TACACAAAAT GAATTGCAA	4860
CCAGACCATT AAGTCGAATA	GTGAGTTCTG AAGGCAGGGC	GTGTTCTCA AACAGTGATG	4920
GCGCGACGGT GTGCGTTAAT	ATCGCTGATA ACGGGCGGT	GCGGTCAGTA ATTGACAAGG	4980
TAGATTCAT CCTGCAATGA	AGTCATTAA TTTCTGTATT	ATTTACTGTG TGGGTTAAAG	5040
TTCAGTACGG GCTTTACCCA	TCTTGTAAAA AATTACGGAG	AATAACAATAA AGTATTAAAG	5100
ACAGGTTATT ATTATG			5116

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1536 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Asn	Lys	Ile	Tyr	Arg	Leu	Lys	Phe	Ser	Lys	Arg	Leu	Asn	Ala	Leu
1				5						10					15

Val	Ala	Val	Ser	Glu	Leu	Ala	Arg	Gly	Cys	Asp	His	Ser	Thr	Glu	Lys
				20			25					30			

Gly	Ser	Glu	Lys	Pro	Ala	Arg	Met	Lys	Val	Arg	His	Leu	Ala	Leu	Lys
				35			40					45			

Pro	Leu	Ser	Ala	Met	Leu	Leu	Ser	Leu	Gly	Val	Thr	Ser	Ile	Pro	Gln
				50			55				60				

Ser	Val	Leu	Ala	Ser	Gly	Leu	Gln	Gly	Met	Asp	Val	Val	His	Gly	Thr
				65		70		75					80		

Ala	Thr	Met	Gln	Val	Asp	Gly	Asn	Lys	Thr	Ile	Ile	Arg	Asn	Ser	Val
				85			90					95			

Asp	Ala	Ile	Ile	Asn	Trp	Lys	Gln	Phe	Asn	Ile	Asp	Gln	Asn	Glu	Met
				100			105					110			

Val	Gln	Phe	Leu	Gln	Glu	Asn	Asn	Asn	Ser	Ala	Val	Phe	Asn	Arg	Val
				115			120					125			

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Ser Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Ile Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Ile Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Ile Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Ile Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ile Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr
 450 455 460
 Ile Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr
 475 480

Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe
 485 490 495
 Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly
 515 520 525
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly
 530 535 540
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn
 545 550 555 560
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp
 565 570 575
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr
 580 585 590
 Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys
 610 615 620
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys
 625 630 635 640
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys
 645 650 655
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu
 660 665 670
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala
 675 680 685
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys
 690 695 700
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile
 705 710 715 720
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser
 725 730 735
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Ser Val Asp Phe Thr
 740 745 750
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn
 755 760 765
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr
 770 775 780
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu
 785 790 795 800
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp
 805 810 815
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu
 820 825 830

Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu
 835 840 845
 Gly Asn Val Thr Ile Asn Asn Asn Ala Asn Val Thr Leu Ile Gly Ser
 850 855 860
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile
 865 870 875 880
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala
 885 890 895
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn
 900 905 910
 Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn
 915 920 925
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser
 930 935 940
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Ser Thr Tyr Arg Thr Ile
 945 950 955 960
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn
 965 970 975
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys
 980 985 990
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln
 995 1000 1005
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala
 1010 1015 1020
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr
 1025 1030 1035 1040
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys
 1045 1050 1055
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr
 1060 1065 1070
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly
 1090 1095 1100
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr
 1105 1110 1115 1120
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Ile Thr Ser His Lys
 1125 1130 1135
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly
 1140 1145 1150
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr
 1155 1160 1165
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu
 1170 1175 1180

Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr
 1185 1190 1195 1200
 Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser
 1205 1210 1215
 Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp
 1220 1225 1230
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu
 1235 1240 1245
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu
 1250 1255 1260
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly
 1265 1270 1275 1280
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn
 1285 1290 1295
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser
 1300 1305 1310
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys
 1315 1320 1325
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile
 1330 1335 1340
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val
 1345 1350 1355 1360
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala
 1365 1370 1375
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val
 1380 1385 1390
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser
 1395 1400 1405
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn
 1410 1415 1420
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys
 1425 1430 1435 1440
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Glu Val
 1445 1450 1455
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu
 1460 1465 1470
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile
 1475 1480 1485
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr
 1490 1495 1500
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser
 1505 1510 1515 1520
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg
 1525 1530 1535

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4937 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAAATATACA	AGATAATAAA	AATAAATCAA	GATTTTGTG	ATGACAAACA	ACAATTACAA	60
CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	AGTATAAATC	CGCCATATAA	120
AATGGTATAA	TCTTCATCT	TTCATCTTAA	ATCTTCATC	TTTCATCTTT	CATTTTCAT	180
CTTTCATCTT	TCATCTTCA	TCTTCATCT	TTCATCTTC	ATCTTCATC	TTTCATCTTT	240
CACATGAAAT	GATGAACCGA	GGGAAGGGAG	GGAGGGCAA	GAATGAAGAG	GGAGCTGAAC	300
GAACGCAAAT	GATAAAGTAA	TTTAATTGTT	CAACTAACCT	TAGGAGAAAA	TATGAACAAG	360
ATATATCGTC	TCAAATTCA	CAAACGCCTG	AATGCTTGG	TTGCTGTGTC	TGAATTGGCA	420
CGGGGTTGTG	ACCATTCCAC	AGAAAAAGGC	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	480
CACTTAGCGT	TAAAGCCACT	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	ATCTATTCCA	540
CAATCTGTTT	TAGCAAGCGG	CTTACAAGGA	ATGGATGTAG	TACACGGCAC	AGCCACTATG	600
CAAGTAGATG	GTAATAAAAC	CATTATCCGC	AACAGTGTG	ACGCTATCAT	TAATTGGAAA	660
CAATTAAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTAC	AAGAAAACAA	CAACTCCGCC	720
GTATTCAACC	GTGTTACATC	TAACCAAATC	TCCAATTAA	AAGGGATTTT	AGATTCTAAC	780
GGACAAGTCT	TTTTAATCAA	CCCAAATGGT	ATCACAATAG	GTAAAGACGC	AATTATTAAC	840
ACTAATGGCT	TTACGGCTTC	TACGCTAGAC	ATTCTAACG	AAAACATCAA	GGCGCGTAAT	900
TTCACCTTCG	AGCAAACCAA	AGATAAAGCG	CTCGCTGAAA	TTGTGAATCA	CGGTTTAATT	960
ACTGTCGGTA	AAGACGGCAG	TGTAATCTT	ATTGGTGGCA	AAGTAAAAAA	CGAGGGTGTG	1020
ATTAGCGTAA	ATGGTGGCAG	CATTCTTTA	CTCGCAGGGC	AAAAAATCAC	CATCAGCGAT	1080
ATAATAAAC	CAACCATTAC	TTACAGCATT	GCCGCGCCTG	AAAATGAAGC	GGTCAATCTG	1140
GGCGATATTT	TTGCCAAAGG	CGGTAACATT	AATGTCCGTG	CTGCCACTAT	TCGAAACCAA	1200
GGTAAACTTT	CTGCTGATTC	TGTAAGCAA	GATAAAAGCG	GCAATATTGT	TCTTCCGCC	1260
AAAGAGGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTCCGCTC	AAAATCAGCA	AGCTAAAGGC	1320
GGCAAGCTGA	TGATTACAGG	CGATAAAGTC	ACATTAACAA	CAGGTGCAGT	TATCGACCTT	1380
TCAGGTAAAG	AAGGGGGAGA	AACTTACCTT	GGCGGTGACG	AGCGCGCGA	AGGTAAAAAC	1440
GGCATTCAAT	TAGCAAAGAA	AACCTCTTAA	AAAAAAGGCT	CAACCATCAA	TGTATCAGGC	1500
AAAGAAAAAG	GCGGACGCGC	TATTGTGTGG	GGCGATATTG	CGTTAATTGA	CGGCAATATT	1560
AACGCTCAAG	GTAGTGGTGA	TATCGCTAAA	ACCGGTGGTT	TTGTGGAGAC	ATCGGGGCAT	1620
TATTATCCA	TTGACAGCAA	TGCAATTGTT	AAAACAAAAG	AGTGGTTGCT	AGACCCTGAT	1680

GATGTAACAA TTGAAGCCGA AGACCCCTT CGCAATAATA CCGGTATAAA TGATGAATT	1740
CCAACAGGCA CCGGTGAAGC AAGCGACCTT AAAAAAAATA GCGAACTCAA AACAAACGCTA	1800
ACCAATACAA CTATTTCAA TTATCTGAAA AACGCCCTGGA CAATGAATAT AACGGCATCA	1860
AGAAAACCTTA CCGTTAATAG CTCAATCAAC ATCGGAAGCA ACTCCCACCTT AATTCTCCAT	1920
AGTAAAGGTC AGCGTGGCGG AGGCCTTCAG ATTGATGGAG ATATTACTTC TAAAGGCGGA	1980
AATTTAACCA TTTATTCTGG CGGATGGTT GATGTTCATA AAAATATTAC GCTTGATCAG	2040
GGTTTTTAA ATATTACCGC CGCTTCCGTA GCTTTGAAG GTGAAATAA CAAAGCACGC	2100
GACGCGGCAA ATGCTAAAAT TGTCGCCAG GGCACGTAA CCATTACAGG AGAGGGAAAA	2160
GATTCAGGG CTAACAACGT ATCTTAAAC GGAACGGGTAA AAGGTCTGAA TATCATTCA	2220
TCAGTGAATA ATTTAACCCAA CAATCTTAGT GGCACAATTA ACATATCTGG GAATATAACA	2280
ATTAACAAA CTACGAGAAA GAACACCTCG TATTGGCAAA CCAGCCATGA TTCCGACTGG	2340
AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAATTTA CCTTTATTAA ATACATTCA	2400
AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG CAGGGTGAA TTTAACGGC	2460
GTAAATGGCA ACATGTCATT CAATCTAAA GAAGGAGCGA AAGTTAATT CAAATTAAAA	2520
CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTG GGTGTTAGC CAATATCACA	2580
GCCACTGGTG GGGGCTCTGT TTTTTTGAT ATATATGCCA ACCATTCTGG CAGAGGGCT	2640
GAGTTAAAAA TGAGTGAAT TAATATCTCT AACGGCGCTA ATTTTACCTT AAATTCCCAT	2700
GTTCGCGGCG ATGACGCTTT TAAAATCAAC AAAGACTTAA CCATAAATGC ACCAACATTCA	2760
AATTCAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG GGTACGCCAG CAATGCCATC	2820
AATTCAACCT ACAACATATC CATTCTGGC GGTAAATGTCA CCCTTGGTGG ACAAAACTCA	2880
AGCAGCAGCA TTACGGGAA TATTACTATC GAGAAAGCAG CAAATGTTAC GCTAGAAGCC	2940
AATAACGCC CTAATCAGCA AAACATAAGG GATAGAGTTA TAAAACCTGG CAGCTTGCTC	3000
GTAAATGGGA GTTTAAGTT AACTGGCGAA AATGCAGATA TTAAAGGCAA TCTCACTATT	3060
TCAGAAAGCG CCACTTTAA AGGAAAGACT AGAGATAACCC TAAATATCAC CGGCAATTTT	3120
ACCAATAATG GCACTGCCGA AATTAATATA ACACAAAGGAG TGGTAAACT TGGCAATGTT	3180
ACCAATGATG GTGATTTAA CATTACCACT CACGCTAAAC GCAACCAAAG AAGCATCATC	3240
GGCGGAGATA TAATCAACAA AAAAGGAAGC TTAAATATTAA CAGACAGTAA TAATGATGCT	3300
GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA ACCTCACGAT TTCTTCCGAT	3360
AAAATTAAATA TCACCAAACA GATAACAATC AAAAAGGGTA TTGATGGAGA GGACTCTAGT	3420
TCAGATGCGA CAAGTAATGC CAACCTAACT ATTAAAACCA AAGAATTGAA ATTGACAGAA	3480
GACCTAAGTA TTTCAGGTTT CAATAAAGCA GAGATTACAG CCAAAGATGG TAGAGATTAA	3540
ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG CAAAACAGT AACTTTAAC	3600
AATGTTAAAG ATTCAAAAAT CTCTGCTGAC GGTCAACATG TGACACTAAA TAGCAAAGTG	3660
AAAACATCTA GCAGCAATGG CGGACGTGAA AGCAATAGCG ACAACGATAC CGGCTTAAC	3720

ATTACTGCAA AAAATGAGA AGTAAACAAA GATATTACTT CTCTCAAAAC AGTAAATATC	3780
ACCGCGTCGG AAAAGGTTAC CACCACAGCA GGCTCGACCA TTAACGCAAC AAATGGCAAA	3840
GCAAGTATTA CAACCAAAAC AGGTGATATC AGCGGTACGA TTTCCGGTAA CACGGTAAGT	3900
GTTAGCGCGA CTGGTGATT AACCCTAAA TCCGGCTCAA AAATTGAAGC GAAATCGGGT	3960
GAGGCTAATG TAACAAGTGC AACAGGTACA ATTGGCGGTAA CAATTCCGG TAATACGGTA	4020
AATGTTACGG CAAACGCTGG CGATTTAACCA GTTGGGAATG GCGCAGAAAT TAATGCGACA	4080
GAAGGAGCTG CAACCTTAAC CGCAACAGGG AATACCTTGA CTACTGAAGC CGGTTCTAGC	4140
ATCACTTCAA CTAAGGGTCA GGTAGACCTC TTGGCTCAGA ATGGTAGCAT CGCAGGAAGC	4200
ATTAATGCTG CTAATGTGAC ATTAAATACT ACAGGCCACCT TAACCACCGT GGCAGGCTCG	4260
GATATTAAAG CAACCAGCGG CACCTGGTT ATTAACGCAA AAGATGCTAA GCTAAATGGT	4320
GATGCATCAG GTGATAGTAC AGAAGTGAAT GCAGTCAACG CAAGCGGCTC TGGTAGTGTG	4380
ACTGCGGCAA CCTCAAGCAG TGTGAATATC ACTGGGGATT TAAACACAGT AAATGGGTTA	4440
AATATCATT CGAAAGATGG TAGAAACACT GTGCGCTTAA GAGGCAAGGA AATTGAGGTG	4500
AAATATATCC AGCCAGGTGT AGCAAGTGTAA GAAGAAGTAA TTGAAGCGAA ACGCGTCCTT	4560
GAAAAAGTAA AAGATTATAC TGATGAAGAA AGAGAAACAT TAGCTAAACT TGGTAGTAAAGT	4620
GCTGTACGTT TTGTTGAGCC AAATAATACA ATTACAGTCA ATACACAAAA TGAATTTACA	4680
ACCAGACCGT CAAGTCAAGT GATAATTCT GAAGGTAAGG CGTGTTCCTC AAGTGGTAAT	4740
GGCGCACGAG TATGTACCAA TGTGCTGAC GATGGACAGC CGTAGTCAGT AATTGACAAG	4800
GTAGATTCA TCCTGCAATG AAGTCATTT ATTTCTGTAT TATTTACTGT GTGGGTTAAA	4860
GTTCAGTACG GGCTTTACCC ATCTGTAAA AAATTACGGA GAATACAATA AAGTATTTTT	4920
AACAGGTTAT TATTATG	4937

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1477 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu	
1 5 10 15	
Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys	
20 25 30	
Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys	
35 40 45	
Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln	
50 55 60	

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
 65 70 75 80
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125
 Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415

Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro
 450 455 460
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys
 465 470 475 480
 Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp
 485 490 495
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile
 500 505 510
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg
 515 520 525
 Gly Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn
 530 535 540
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr
 545 550 555 560
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu
 565 570 575
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala
 580 585 590
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn
 595 600 605
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser
 610 615 620
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly
 625 630 635 640
 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln
 645 650 655
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr
 660 665 670
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly
 675 680 685
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val
 690 695 700
 Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe
 705 710 715 720
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile
 725 730 735
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Gly Ser Val Phe Phe
 740 745 750
 Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser
 755 760 765

Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val
 770 775 780
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala
 785 790 795 800
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp
 805 810 815
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu
 820 825 830
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ser Ile Thr
 835 840 845
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn
 850 855 860
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly
 865 870 875 880
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp
 885 890 895
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys
 900 905 910
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr
 915 920 925
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr
 930 935 940
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg
 945 950 955 960
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile
 965 970 975
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr
 995 1000 1005
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser
 1010 1015 1020
 Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn
 1060 1065 1070
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser
 1075 1080 1085
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys
 1090 1095 1100
 Thr Ser Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr
 1105 1110 1115 1120

Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr
 1125 1130 1135
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr
 1140 1145 1150
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr
 1155 1160 1165
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val
 1170 1175 1180
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala
 1185 1190 1195 1200
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly
 1205 1210 1215
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu
 1220 1225 1230
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr
 1235 1240 1245
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile
 1250 1255 1260
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile
 1265 1270 1275 1280
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr
 1285 1290 1295
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu
 1300 1305 1310
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp
 1315 1320 1325
 Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr
 1330 1335 1340
 Ala Ala Thr Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val
 1345 1350 1355 1360
 Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu
 1365 1370 1375
 Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser
 1380 1385 1390
 Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp
 1395 1400 1405
 Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala
 1410 1415 1420
 Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn
 1425 1430 1435 1440
 Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys
 1445 1450 1455
 Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala
 1460 1465 1470

Asp Asp Gly Gln Pro
1475

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9171 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTT GCAGTCTATA TGCAAATATT TTAAAAAATA GTATAAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTCATC TCTTCATCTT TTCACTTTTC ATCTTCATC	180
TTTCATCTTT CATCTTCAT CTTTCATCTT TCATCTTCATC TCTTCATCTT TTCACTTTTC	240
ACATGAAATG ATGAACCGAG GGAAGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG	300
AACGCAAATG ATAAAGTAAT TTAATTGTTCA AACTAACCTT AGGAGAAAAT ATGAACAAGA	360
TATATCGTCT CAAATTCTAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT GAATTGGCAC	420
GGGGTTGTGA CCATTCCACA GAAAAAGGCA GCGAAAACC TGCTCCATG AAAGTGCCTC	480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTT AGGTGTAACA TCTATTCCAC	540
AATCTGTTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC	600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTGTA CGCTATCATT AATTGGAAAC	660
AATTAAACAT CGACCAAAAT GAAATGGTGC AGTTTTACA AGAAAACAAC AACTCCGCCG	720
TATTCAACCG TGTTACATCT AACCAAATCT CCCAATTAAA AGGGATTTA GATTCTAACG	780
GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA	840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG GCGCGTAATT	900
TCACCTTCGA GCAAACCAAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC GGTTTAATT	960
CTGTCGGTAA AGACGGCACT GTAAATCTT TTGGTGGCAA AGTGAAAAC GAGGGTGTGA	1020
TTAGCGTAA TGTTGGCAGC ATTTCTTAC TCGCAGGGCA AAAAATCACC ATCAGCGATA	1080
TAATAAAACC AACCATTACT TACAGCATTG CCGCGCCTGA AAATGAAGCG GTCAATCTGG	1140
GCGATATTTT TGCCAAAGGC GGTAACATTA ATGTCCGTGC TGCCACTATT CGAAACCAAG	1200
CTTCCGCCA AAGAGGGTGA AGCGGAAATT GGCAGGTGAA TTTCCGCTCA AAATCAGCAA	1260
GCTAAAGGCG GCAAGCTGAT GATTACAGGC GATAAAAGTCA CATTAAAAAC AGGTGCAGTT	1320
ATCGACCTTT CAGGTAAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGGGCGAA	1380
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCTCTTTAG AAAAAGGCTC AACCATCAAT	1440
GTATCAGGCA AAGAAAAAGG CGGACCGCCT ATTGTGTGGG GCGATATTGC GTTAATTGAC	1500

GGCAATATTA ACGCTCAAGG TAGTGGTGAT ATCGCTAAAA CCGGTGGTTT TGTGGAGACG	1560
TCGGGGCATG ATTTATTCAT CAAAGACAAT GCAATTGTTG ACGCCAAAGA GTGGTTGTTA	1620
GACCCGGATA ATGTATCTAT TAATGCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC	1680
GATGAATACA CGGGATCCGG GAATAGTGCC AGCACCCAA AACGAAACAA AGAAAAGACA	1740
ACATTAACAA ACACAACTCT TGAGAGTATA CTAAAAAAAG GTACCTTGT TAACATCACT	1800
GCTAATCAAC GCATCTATGT CAATAGCTCC ATTAATTAT CCAATGGCAG CTTAACTCTT	1860
TGGAGTGAGG GTCGGAGCGG TGGCGCGTT GAGATTAACA ACGATATTAC CACCGGTGAT	1920
GATACCAGAG GTGCAAACCTT AACAAATTAC TCAGGCGGCT GGGTTGATGT TCATAAAAAT	1980
ATCTCACTCG GGGCGCAAGG TAACATAAAC ATTACAGCTA AACAGATAT CGCCTTTGAG	2040
AAAGGAAGCA ACCAAGTCAT TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAAGGT	2100
TTTAGATTTA ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT CACCACTAAA	2160
AGAACCAATA AATACGCTAT CACAAATAAA TTTGAAGGGA CTTAAATAT TTCAGGGAAA	2220
GTGAACATCT CAATGGTTT ACCTAAAAAT GAAAGTGGAT ATGATAAATT CAAAGGACGC	2280
ACTTACTGGA ATTTAACCTC GAAAGTGGAT ATGATAAATT CAAAGGACGC CCTCACTATT	2340
GACTCCAGAG GAAGCGATAG TGCAGGCACA CTTACCCAGC CTTATAATT AAACGGTATA	2400
TCATTCAACA AAGACACTAC CTTTAATGTT GAACGAAATG CAAGAGTCAA CTTTGACATC	2460
AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGAATT ACGCATCATT TAATGGAAAC	2520
ATTCAGTTT CGGGAGGGGG GAGTGGTGAT TTCACACTTC TCGCCTCATC CTCTAACGTC	2580
CAAACCCCG GTGTAGTTAT AAATTCTAAA TACTTTAATG TTTCAACAGG GTCAAGTTA	2640
AGATTTAAAA CTTCAAGGCTC AACAAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA	2700
AATGCCACCG GAGGCAACAT AACACTTTG CAAGTGAAG GCACCGATGG AATGATTGGT	2760
AAAGGCATTG TAGCCAAAAA AAACATAACC TTTGAAGGAG GTAAGATGAG GTTGGCTCC	2820
AGGAAAGCCG TAACAGAAAT CGAAGGCAAT GTTACTATCA ATAACAAACGC TAACGTCACT	2880
CTTATCGGTT CGGATTTGA CAACCATCAA AAACCTTTAA CTATTAAAAA AGATGTCATC	2940
ATTAATAGCG GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC	3000
GTTGAAAGTA ACGCTAATT TCAAAGCTATC ACAAAATTCA CTTTTAATGT AGGCGGCTTG	3060
TTTGACAACA AAGGCAATTG AAATATTTC ATTGCCAAG GAGGGCTCG CTTTAAAGAC	3120
ATTGATAATT CCAAGAATT AAGCATCACC ACCAACTCCA GCTCCACTTA CCGCACTATT	3180
ATAAGCGGCA ATATAACCAA TAAAAACGGT GATTAAATA TTACGAACGA AGGTAGTGAT	3240
ACTGAAATGC AAATTGGCGG CGATGTCTCG CAAAAAGAAG GTAATCTCAC GATTCTTCT	3300
GACAAAATCA ATATTACCAA ACAGATAACA ATCAAGGCAG GTGTTGATGG GGAGAATTCC	3360
GATTCAAGACG CGACAAACAA TGCCAATCTA ACCATTAAAA CCAAAGAATT GAAATTAACG	3420
CAAGACCTAA ATATTCAGG TTTCAATAAA GCAGAGATTA CAGCTAAAGA TGGTAGTGAT	3480
TTAACTATTG GTAACACCAA TAGTGCTGAT GGTACTAATG CCAAAAAAGT AACCTTTAAC	3540

CAGGTTAAAG ATTCAAAAAT CTCTGCTGAC GGTCACAAGG TGACACTACA CAGCAAAGTG	3600
GAAACATCCG GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC CGGCTTAACT	3660
ATCGATGCAA AAAATGTAAC AGTAAACAAAC AATATTACTT CTCACAAAGC AGTGAGCATC	3720
TCTGCGACAA GTGGAGAAAT TACCACTAAA ACAGGTACAA CCATTAACGC AACCACTGGT	3780
AACGTGGAGA TAACCGCTCA AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC	3840
TCTGTAACAC TTACTGCAAC CGAGGGCGCT CTTGCTGTAA GCAATATTTC GGGCAACACC	3900
GTTACTGTTA CTGCAAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC AATTAAAGGA	3960
ACCGAGAGTG TAACCACTTC AAGTCATCA GGCGATATCG GCGGTACGAT TTCTGGTGGC	4020
ACAGTAGAGG TTAAAGCAAC CGAAAGTTA ACCACTCAAT CCAATTCAA AATTAAAGCA	4080
ACAAACAGGCG AGGCTAACGT AACAAAGTGCA ACAGGTACAA TTGGTGGTAC GATTTCGGT	4140
AATACGGTAA ATGTTACGGC AAACGCTGGC GATTTAACAG TTGGGAATGG CGCAGAAATT	4200
AATGCGACAG AAGGAGCTGC AACCTTAAC ACTCATCGG GCAAATTAAC TACCGAAGCT	4260
AGTTCACACA TTACTTCAGC CAAGGGTCAG GTAAATCTTT CAGCTCAGGA TGTTAGCGTT	4320
GCAGGAAGTA TTAATGCCGC CAATGTGACA CTAAATACTA CAGGCACCTT AACTACCGTG	4380
AAGGGTTCAA ACATTAATGC AACCAAGCGGT ACCTTGGTTA TTAACGCAA AGACGCTGAG	4440
CTAAATGGCG CAGCATTGGG TAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC	4500
GGCAGCGTAA TCGCGACAAC CTCAAGCAGA GTGAACATCA CTGGGGATTT AATCACAATA	4560
AATGGATTAA ATATCATTTC AAAAAACGGT ATAAACACCG TACTGTAAA AGCGTTAAA	4620
ATTGATGTGA AATACATTCA ACCGGGTATA GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA	4680
CGCATCCTTG AGAAGGTAAA AGATTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT	4740
GGCGTAAGTG CTGTACGTTT TATTGAGCCA ATAATACAA TTACAGTCGA TACACAAAAT	4800
GAATTTGCAA CCAGACCATT AAGTCGAATA GTGATTCTG AAGGCAGGGC GTGTTCTCA	4860
AACAGTGATG GCGCGACGGT GTGCGTTAAT ATCGCTGATA ACGGGCGGT GCGGTAGTA	4920
ATTGACAAGG TAGATTTCAT CCTGCAATGA AGTCATTTTA TTTTCGTATT ATTTACTGTG	4980
TGGGTTAAAG TTCAGTACGG GCTTACCCA TCTTGTAAAA AATTACGGAG AATACAATAA	5040
AGTATTTTA ACAGGTTATT ATTATGAAA ATATAAAAAG CAGATTAAC CTCAGTGCAA	5100
TATCAGTATT GCTTGGCCTG GCTTCTTCAT CATTGTATGC AGAAGAAGCG TTTTAGTAA	5160
AAGGCTTCA GTTATCTGGT GCACTTGAAA CTTTAAGTGA AGACGCCAA CTGTCTGTAG	5220
CAAAATCTTT ATCTAAATAC CAAGGCTCGC AAACTTAAC AAACCTAAAA ACAGCACAGC	5280
TTGAATTACA GGCTGTGCTA GATAAGATTG AGCCAAATAA GTTTGATGTG ATATTGCCAC	5340
AACAAACCAT TACGGATGGC AATATTATGT TTGAGCTAGT CTCGAAATCA GCCGCAGAAA	5400
GCCAAGTTTT TTATAAGGCG AGCCAGGGTT ATAGTGAAGA AAATATCGCT CGTAGCCTGC	5460
CATCTTGAA ACAAGGAAAA GTGTATGAAG ATGGTCGTCA GTGGTTCGAT TTGCGTGAAT	5520
TCAATATGGC AAAAGAAAAAT CCACTAAAG TCACTCGCGT GCATTACGAG TTAAACCCCTA	5580

AAAACAAAAC	CTCTGATTG	GTAGTTGCAG	GTTCGAG	TTTTGGCAAA	ACCGTAGCT	5640
TTGTTTCCTA	TGATAATTTC	GGCGCAAGGG	AGTTAACTA	TCAACGTGTA	AGTCTAGGTT	5700
TTGTAATGC	CAATTTGACC	GGACATGATG	ATGTATTAAA	TCTAAACGCA	TTGACCAATG	5760
TAAAAGCACC	ATCAAAATCT	TATGCGGTAG	GCATAGGATA	TACTTATCCG	TTTTATGATA	5820
AACACCAATC	CTTAAGTCTT	TATACCAGCA	TGAGTTATGC	TGATTCTAAT	GATATCGACG	5880
GCTTACCAAG	TGCGATTAAT	CGTAAATTAT	CAAAAGGTCA	ATCTATCTCT	GCGAATCTGA	5940
AATGGAGTTA	TTATCTCCCG	ACATTTAACCC	TTGGAATGGA	AGACCAGTTT	AAAATTAATT	6000
TAGGCTACAA	CTACCGCCAT	ATTAATCAAA	CATCCGAGTT	AAACACCCCTG	GGTGCACACGA	6060
AGAAAAAAATT	TGCAGTATCA	GGCGTAAGTG	CAGGCATTGA	TGGACATATC	CAATTTACCC	6120
CTAAAACAAT	CTTTAATATT	GATTAACTC	ATCATTATTA	CGCGAGTAAA	TTACCAGGCT	6180
CTTTTGAAT	GGAGCGCATT	GGCGAACAT	TTAATCGCAG	CTATCACATT	AGCACAGCCA	6240
GTTCAGGGTT	GAGTCAAGAG	TTTGCTCAAG	GTTGGCATT	TAGCAGTCAA	TTATCGGGTC	6300
AGTTTACTCT	ACAAGATATA	AGTAGCATAG	ATTTATTCTC	TGTAACAGGT	ACTTATGGCG	6360
TCAGAGGCTT	TAAATACGGC	GGTGCAGTG	GTGAGCGCGG	TCTTGTATGG	CGTAATGAAT	6420
TAAGTATGCC	AAAATACACC	CGCTTCAAA	TCAGCCCTTA	TGCGTTTTAT	GATGCAGGTC	6480
AGTCCCGTTA	TAATAGCGAA	AATGCTAAAA	CTTACGGCGA	AGATATGCAC	ACGGTATCCT	6540
CTGCGGGTTT	AGGCATTAAA	ACCTCTCCTA	CACAAAACCTT	AAGCTTAGAT	GCTTTGTTG	6600
CTCGTCGCTT	TGCAAATGCC	AATAGTGACA	ATTTGAATGG	CAACAAAAAA	CGCACAAAGCT	6660
CACCTACAAC	CTTCTGGGTT	AGATTAACAT	TCAGTTCTA	ACCCGAAAT	TTAATCAACT	6720
GGTAAGCGTT	CCGCCTACCA	GTTTATACT	ATATGCTTTA	CCCGCCAATT	TACAGTCTAT	6780
ACGCAACCCT	GTTCATCC	TTATATATCA	AAACAACTAA	GCAAACCAAG	CAAACCAAGC	6840
AAACCAAGCA	AACCAAGCAA	ACCAAGCAA	CCAAGCAAAC	CAAGCAAACC	AAGCAAACCA	6900
AGCAAACCAA	GCAAACCAAG	CAAACCAAGC	AAACCAAGCA	ATGCTAAAAA	ACAATTTATA	6960
TGATAAACTA	AAACATACTC	CATACCATGG	CAATACAAGG	GATTAAATAA	TATGACAAAA	7020
GAAAATTTAC	AAAGTGTCC	ACAAAATACG	ACCGCTTCAC	TTGTAGAATC	AAACAACGAC	7080
CAAACCTCCC	TGCAAATACT	AAACAAACCA	CCCAAACCCA	ACCTATTACG	CCTGGAACAA	7140
CATGTCGCCA	AAAAAGATTA	TGAGCTTGCT	TGCCGCGAAT	TAATGGCGAT	TTTGGAAAAA	7200
ATGGACGCTA	ATTTGGAGG	CGTTCACGAT	ATTGAATTG	ACGCACCTGC	TCAGCTGGCA	7260
TATCTACCCG	AAAAACTACT	AATTCACTT	GCCACTCGTC	TCGCTAATGC	AATTACAACA	7320
CTCTTTCCG	ACCCCGAATT	GGCAATTCC	GAAGAAGGGG	CATTAAAGAT	GATTAGCCTG	7380
CAACGCTGGT	TGACGCTGAT	TTTGCCCTCT	TCCCCCTACG	TTAACGCAGA	CCATATTCTC	7440
AATAAAATATA	ATATCAACCC	AGATTCGAA	GGTGGCTTC	ATTTAGCAAC	AGACAACTCT	7500
TCTATTGCTA	AATTCTGTAT	TTTTTACTTA	CCCGAATCCA	ATGTCAATAT	GAGTTTAGAT	7560
GCCTTATGGG	CAGGAATCA	ACAACTTGT	GCTTCATTGT	GTTCAGTCTTCA		7620

CGTTTTATTG GTACTGCATC TGCCTTCAT AAAAGAGCGG TGGTTTACA GTGGTTTCCT	7680
AAAAAACTCG CCGAAATTGC TAATTTAGAT GAATTGCCTG CAAATATCCT TCATGATGTA	7740
TATATGCACT GCAGTTATGA TTTAGAAAA AACAAAGCACG ATGTTAAGCG TCCATTAAAC	7800
GAACTTGTCC GCAAGCATAT CCTCACGCAA GGATGGCAAG ACCGCTACCT TTACACCTTA	7860
GGTAAAAAGG ACGGAAACC TGTGATGATG GTACTGCTTG AACATTTAA TTCGGGACAT	7920
TCGATTTATC GCACGCATTC AACTTCAATG ATTGCTGCTC GAGAAAAATT CTATTTAGTC	7980
GGCTTAGGCC ATGAGGGCGT TGATAACATA GGTGAGAAG TGTTGACGA GTTCTTGAA	8040
ATCAGTAGCA ATAATATAAT GGAGAGACTG TTTTTATCC GTAAACAGTG CGAAACTTTC	8100
CAACCCGCAG TGTCTATAT GCCAAGCATT GGCACTGGATA TTACACGAT TTTGTGAGC	8160
AACACTCGGC TTGCCCCAT TCAAGCTGTA GCCTTGGTC ATCCIGCCAC TACGCATTCT	8220
GAATTATTG ATTATGTCAT CGTAGAAGAT GATTATGTGG GCAGTGAAGA TTGTTTTAGC	8280
GAAACCCTT TAGGCTTACC CAAAGATGCC CTACCTTATG TACCATCTGC ACTCGCCCCA	8340
CAAAAAGTGG ATTATGTA CAGGGAAAAC CCTGAAGTAG TCAATATCGG TATTGCCGCT	8400
ACCACAAATGA AATTAAACCC TGAATTTTG CTAACATTGC AAGAAATCAG AGATAAAAGCT	8460
AAAGTCAAA TACATTTCA TTTCGCACTT GGACAATCAA CAGGCTTGAC ACACCCTTAT	8520
GTCAAATGGT TTATCGAAAG CTATTTAGGT GACGATGCCA CTGCACATCC CCACGCACCT	8580
TATCACGATT ATCTGGCAAT ATTGCGTGAT TGCGATATGC TACTAAATCC GTTCCCTTC	8640
GGTAATACTA ACGGCATAAT TGATATGGTT ACATTAGGTT TAGTTGGTGT ATGAAACACG	8700
GGGGATGAAG TACATGAACA TATTGATGAA GGTCTGTTA AACGCTTAGG ACTACCAGAA	8760
TGGCTGATAG CCGACACACG AGAAACATAT ATTGAATGTG CTTTGCCTCT AGCAGAAAAC	8820
CATCAAGAAC GCCTTGAAC CCGTCGTTAC ATCATAGAAA ACAACGGCTT ACAAAAGCTT	8880
TTTACAGGCG ACCCTCGTCC ATTGGCAAA ATACTGCTTA AGAAAACAAA TGAATGGAAG	8940
CGGAAGCACT TGAGTAAAAA ATAACGGTTT TTTAAAGTAA AAGTGCAGTT AATTTTCAA	9000
CGCTTTAAA AACCTCTCAA AAATCAACCG CACTTTATC TTTATAACGC TCCCGCGCGC	9060
TGACAGTTA TCTCTTCTT AAAATACCCA TAAAATTGTG GCAATAGTT GGTAATCAA	9120
TTCAATTGTT GATACGGCAA ACTAAAGACG GCGCGTTCTT CGGCAGTCAT C	9171

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9323 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCCACTTCA ATTTTGGATT GTTGAATTC AACTAACCAA AAAGTGCAGTT TAAAATCTGT

GGAGAAAATA	GGTTGTAGTG	AAGAACGAGG	TAATTGTTCA	AAAGGATAAA	GCTCTCTTAA	120
TTGGGCATTG	GTTGGCGTTT	CTTTTCGGT	TAATAGTAAA	TTATATTCTG	GACGACTATG	180
CAATCCACCA	ACAACTTAC	CGTTGGTTT	AAGCGTTAAT	GTAAGTTCTT	GCTCTCTTG	240
GCGAATAACGT	AATCCCATT	TTTGTGTTAGC	AAGAAAATGA	TCGGGATAAT	CATAATAGGT	300
GTTGCCAAA	AATAAATTTT	GATGTTCTAA	AATCATAAAAT	TTTGCAAGAT	ATTGTGGCAA	360
TTCAATAACCT	ATTTGTGGCG	AAATGCCAA	TTTTAATTCA	ATTTCTGT	GCATAATATT	420
TCCCACCTCAA	ATCAACTGGT	TAATATACA	AGATAATAAA	AATAAATCAA	GATTTTGTG	480
ATGACAAACA	ACAATTACAA	CACCTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	540
AGTATAAATC	CGCCATATAA	AATGGTATAA	TCTTCATCT	TTCATCTTC	ATCTTCATC	600
TTTCATCTTT	CATCTTCAT	CTTTCATCTT	TCATCTTC	TCTTCATCT	TTCATCTTC	660
ATCTTCATC	TTTCATCTTT	CACATGAAAT	GATGAACCGA	GGGAAGGGAG	GGAGGGCAA	720
GAATGAAGAG	GGAGCTGAAC	GAACGAAAT	GATAAAAGTAA	TTTAATTGTT	CAACTAACCT	780
TAGGAGAAAA	TATGAACAAG	ATATATCGTC	TCAAATTCA	CAAACGCCTG	AATGCTTTGG	840
TTGCTGTGTC	TGAATTGGCA	CGGGGTTGTG	ACCATTCCAC	AGAAAAAGGC	AGCGAAAAAC	900
CTGCTCGCAT	GAAAGTGCCT	CACTTAGCGT	TAAAGCCACT	TTCCGCTATG	TTACTATCTT	960
TAGGTGTAAC	ATCTATTCCA	CAATCTGTTT	TAGCAAGCGG	CAATTAAACA	TCGACCAAAA	1020
TGAAATGGTG	CAGTTTTAC	AAGAAAACAA	GTAATAAAAC	CATTATCCGC	AACAGTGGT	1080
ACGCTATCAT	TAATTGGAA	CAATTAAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTAC	1140
AAGAAAACAA	CAACTCCGCC	GTATTCAACC	GTGTTACATC	TAACCAAATC	TCCCAATTAA	1200
AAGGGATTT	AGATTCTAAC	GGACAAGTCT	TTTTAATCAA	CCCAAATGGT	ATCACAATAG	1260
GTAAAGACGC	AATTATTAAC	ACTAATGGCT	TTACGGCTTC	TACGCTAGAC	ATTTCTAACG	1320
AAAACATCAA	GGCGCGTAAT	TTCACCTTCG	AGCAAACCAA	AGATAAAAGCG	CTCGCTGAAA	1380
TTGTGAATCA	CGGTTTAATT	ACTGTCGGTA	AAGACGGCAG	TGTAATCTT	ATTGGTGGCA	1440
AAGTGAAAAA	CGAGGGTGTG	ATTAGCGTAA	ATGGTGGCAG	CATTCTTTA	CTCGCAGGGC	1500
AAAAAATCAC	CATCAGCGAT	ATAATAAAC	CAACCATTAC	TTACAGCATT	GCCGCGCCTG	1560
AAAATGAAGC	GGTCAATCTG	GGCGATATT	TTGCCAAAGG	CGGTAACATT	AATGTCCGTG	1620
CTGCCACTAT	TCGAAACCAA	GGTAAACTTT	CTGCTGATT	TGTAAGCAA	GATAAAAGCG	1680
GCAATATTGT	TCTTCCGCC	AAAGAGGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTTCCGCTC	1740
AAAATCAGCA	AGCTAAAGGC	GGCAAGCTGA	TGATAAAAGTC	CGATAAAAGTC	ACATTAACAA	1800
CAGGTGCAGT	TATCGACCTT	TCAGGTAAAG	AAGGGGGAGA	AACTTACCTT	GGCGGTGACG	1860
AGCGCGGCCGA	AGGTAAAAAC	GGCATTCAAT	TAGCAAAGAA	AACCTCTTTA	AAAAAAGGCT	1920
CAACCATCAA	TGTATCAGGC	AAAGAAAAAG	CGGGACGCGC	TATTGTGTGG	GGCGATATTG	1980
CGTTAATTGA	CGGCAATATT	AACGCTCAAG	GTAGTGGTGA	TATCGCTAAA	ACCGGTGGTT	2040
TTGTGGAGAC	ATCGGGGCAT	TATTTATCCA	TTGACAGCAA	TGCAATTGTT	AAAACAAAAG	2100

AGTGGTTGCT AGACCCTGAT GATGTAACAA TTGAAGCCGA AGACCCCTT CGCAATAATA	2160
CCGGTATAAA TGATGAATT CCAACAGGCA CCGGTGAAGC AAGCGACCCT AAAAAAAATA	2220
GCGAACTCAA AACAAACGCTA ACCAATACAA CTATTCAAA TTATCTGAAA AACGCCTGGA	2280
CAATGAATAT AACGGCATCA AGAAAACCTTA CCGTTAATAG CTCATCACAC ATCGGAAGCA	2340
ACTCCCACCTT AATTCTCCAT AGTAAAGGTC AGCGTGGCGG AGGCCTTCAG ATTGATGGAG	2400
ATATTACTTC TAAAGGCGGA AATTAAACCA TTTATTCTGG CGGATGGGTT GATGTTCATATA	2460
AAAATATTAC GCTTGATCAG GGTTTTTAA ATATTACCGC CGCTTCCGTA GCTTTGAAG	2520
GTGGAAATAA CAAAGCACGC GACGCGCAA ATGCTAAAAT TGTCGCCAG GGCAGTGTAA	2580
CCATTACAGG AGAGGGAAAA GATTCAGGG CTAACAACGT ATCTTAAAC GGAACGGGTA	2640
AAGGTCTGAA TATCATTCA TCAGTGAATA ATTTAACCA CAATCTTAGT GGCACAATTAA	2700
ACATATCTGG GAATATAACA ATTAACAAA CTACGAGAAA GAACACCTCG TATTGGCAA	2760
CCAGCCATGA TTGCACTGG AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAATTTA	2820
CCTTTATTAA ATACATTCA AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG	2880
CAGGGTGAA TTTAACGGC GTAAATGGCA ACATGTCATT CAATCTAAA GAAGGAGCGA	2940
AAGTTAATT CAAATTAAAA CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTCA	3000
GGTTTTAGC CAATATCACA GCCACTGGTG GGGCTCTGT TTTTTTGAT ATATATGCCA	3060
ACCATTCTGG CAGAGGGGCT GAGTTAAAAA TGAGTGAAT TAATATCTCT AACGGCGCTA	3120
ATTTTACCTT AAATTCCCAT GTTCGCGCG ATGACGCTTT TAAAATCAAC AAAGACTTAA	3180
CCATAAAATGC AACCAATTCA AATTCAAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG	3240
GGTACGCACG CAATGCCATC AATTCAACCT ACAACATATC CATTCTGGC GGTAAATGTCA	3300
CCCTTGGTGG ACAAAACTCA AGCAGCAGCA TTACGGGAA TATTACTATC GAGAAAGCAG	3360
CAAATGTTAC GCTAGAAGCC AATAACGCC CTAATCAGCA AAACATAAGG GATAGAGTTA	3420
TAAAACTTGG CAGCTTGCTC GTTAATGGGA GTTTAAGTTT AACTGGCGAA AATGCAGATA	3480
TTAAAGGCAA TCTCACTATT TCAGAAAGCG CCACTTTAA AGGAAAGACT AGAGATAACCC	3540
TAAATATCAC CGGCAATTTC ACCAATAATG GCACTGCCGA AATTAATATA ACACAAGGAG	3600
TGGTAAACT TGGCAATGTT ACCAATGATG GTGATTTAAA CATTACCACT CACGCTAAC	3660
GCAACCAAAG AAGCATCATC GGCGGAGATA TAATCAACAA AAAAGGAAGC TTAAATATTA	3720
CAGACAGTAA TAATGATGCT GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA	3780
ACCTCACGAT TTCTTCCGAT AAAATTAAATA TCACCAAACA GATAACAATC AAAAGGGTA	3840
TTGATGGAGA GGACTCTAGT TCAGATGCCA CAAGTAATGC CAACCTAACT ATTAAAACCA	3900
AAGAATTGAA ATTGACAGAA GACCTAAGTA TTTCAGGTTT CAATAAAGCA GAGATTACAG	3960
CCAAAGATGG TAGAGATTAA ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG	4020
CCAAAACAGT AACTTTAAC AATGTTAAAG ATTCAAAAT CTCTGCTGAC GGTCACAATG	4080
TGACACTAAA TAGCAAAGTG AAAACATCTA GCAGCAATGG CGGACGTGAA AGCAATAGCG	4140

ACAACGATA	CGGCTTA	ACT ATTACTGCAA	AAAATGTAGA	AGTAAACAAA	GATATTACTT	4200		
CTCTCAAAAC	AGTAAATATC	ACCGCGTCGG	AAAAGGTTAC	CACCACAGCA	GGCTCGACCA	4260		
TTAACGCAAC	AAATGGCAA	GCAAGTATT	CAACCAAAAC	AGGTGATATC	AGCGGTACGA	4320		
TTTCCGGTAA	CACGGTAAGT	GTTAGCGCGA	CTGGTGATT	AACCAC	AA TCCGGCTCAA	4380		
AAATTGAAGC	GAAATCGGGT	GAGGCTAATG	TAACAAGTGC	AACAGGTACA	ATTGGCGGT	4440		
CAATTTCGG	TAATACGGTA	AATGTTACGG	CAAACGCTGG	CGATTTAAC	GTTGGGAATG	4500		
GCGCAGAAAT	TAATGCGACA	GAAGGAGCTG	CAACCTAAC	CGCAACAGGG	AATACCTTGA	4560		
CTACTGAAGC	CGGTTCTAGC	ATCACTTCAA	CTAAGGGTCA	GGTAGACCTC	TTGGCTCAGA	4620		
ATGGTAGCAT	CGCAGGAAGC	ATTAATGCTG	CTAATGTGAC	ATTAAATACT	ACAGGCACCT	4680		
TAACCACCGT	GGCAGGCTCG	GATATTAAAG	CAACCAGCGG	CACCTTGGTT	ATTAACGCAA	4740		
AAGATGCTAA	GCTAAATGGT	GATGCATCAG	GTGATAGTAC	AGAAGTGAAT	GCAGTCACG	4800		
ACTGGGGATT	TGGTAGTGTG	ACTGCGGCAA	CCTCAAGCAG	TGTGAATATC	ACTGGGGATT	4860		
TAAACACAGT	AAATGGGTTA	AATATCATT	CGAAAGATGG	TAGAAACACT	GTGCGCTTAA	4920		
GAGGCAAGGA	AATTGAGGTG	AAATATATCC	AGCCAGGTGT	AGCAAGTGT	GAAGAAGTAA	4980		
TTGAAGCGAA	ACGCGTCCTT	GAAAAAGTAA	AAGATTTATC	TGATGAAGAA	AGAGAAACAT	5040		
TAGCTAAACT	TGGTAGTAA	GCTGTACGTT	TTGTTGAGCC	AAATAATACA	ATTACAGTCA	5100		
ATACACAAAA	TGAATTTACA	ACCAGACCGT	CAAGTCAGT	GATAATTCT	GAAGGTAAGG	5160		
CGTGTTCCTC	AAAGTGGTAAT	GGCGCACGAG	TATGTACCAA	TGTTGCTGAC	GATGGACAGC	5220		
CGTAGTCAGT	AATTGACAAG	GTAGATTCA	TCCTGCAATG	AAGTCATT	TTTTCGTAT	5280		
TATTTACTGT	GTGGGTTAAA	GTTCAGTACG	GGCTTACCC	ATCTGTAAA	AAATTACGGA	5340		
GAATACAATA	AAAGTATTTT	AAACAGGTTAT	TATTATGAA	AAATAAAAAA	GCAGATTAAA	5400		
ACTCAGTGCA	ATATCAGTAT	TGCTTGGCCT	GGCTTCTTCA	TCATTGTATG	CAGAAGAACG	5460		
GTTTTTAGTA	AAAGGCTTTC	AGTTATCTGG	TGCAC	TTAAGTG	AAGACGCCA	5520		
ACTGTCTGTA	GC	AAAATCTT	TATCTAAATA	CCAAGGCTCG	CAAAC	TTAA	CAAACCTAAA	5580
AACAGCACAG	CTTGAATTAC	AGGCTGTGCT	AGATAAGATT	GAGC	AAATA	AATTGATGT	5640	
GATATTGCCG	CAACAAACCA	TTACGGATGG	CAATATCATG	TTTGAGCTAG	TCTCGAAATC	5700		
AGCCGCAGAA	AGCCAAGTTT	TTTATAAGGC	GAGCCAGGGT	TATAGTGAAG	AAAATATCGC	5760		
TCGTAGCCTG	CCATCTTGA	AAACAGGAAA	AGTGTATGAA	GATGGTCGTC	AGTGGTTCGA	5820		
TTTGCCTGAA	TTTAATATGG	CAAAAGAAAA	CCCGCTTAAG	GTTACCCGTG	TACATTACGA	5880		
ACTAAACCC	TTTAAACAAA	CCTCTAA	TTT	GATAATTGCG	GGCTTCTCGC	CTTTTGGTAA	5940	
AACCGTAGC	TTTATTTCTT	ATGATAATT	CGCGCGAGA	GAGTTAA	ACTAACGTGT	6000		
AAGCTTGGGT	TTTGTAA	CTTAAAC	TGGTCATGAT	GATGTGTAA	TTATACCA	6060		
ATGAGTTATG	CTGATTCTAA	TGATATCGAC	GGCTTACCAA	GTGCGATTAA	TCGTAAATT	6120		
TCAAAAGGTC	AATCTATCTC	TGCGAATCTG	AAATGGAGTT	ATTATCTCCC	AACATTTAAC	6180		

CTTGGCATGG AAGACCAATT TAAAATTAAT TTAGGCTACA ACTACCGCCA TATTAATCAA	6240
ACCTCCGCGT TAAATCGCTT GGGTGAAACG AAGAAAAAAT TTGCAGTATC AGGCGTAAGT	6300
GCAGGCATTG ATGGACATAT CCAATTACCC CCTAAAACAA TCTTTAATAT TGATTTAACT	6360
CATCATTATT ACGCGAGTAA ATTACCAAGGC TCTTTGGAA TGGAGCGCAT TGGCGAAACA	6420
TTTAATCGCA GCTATCACAT TAGCACAGCC AGTTTAGGGT TGAGTCAGA GTTTGCTCAA	6480
GGTTGGCATT TTAGCAGTCA ATTATCAGGT CAATTTACTC TACAAGATAT TAGCAGTATA	6540
GATTATTCT CTGTAACAGG TACTTATGGC GTCAGAGGCT TTAAATACGG CGGTGCAAGT	6600
GGTGAGCGCG GTCTTGTATG GCGTAATGAA TTAAGTATGC CAAAATACAC CCGCTTCCAA	6660
ATCAGCCCTT ATGCGTTTA TGATGCAGGT CAGTTCCGTT ATAATAGCGA AAATGCTAAA	6720
ACTTACGGCG AAGATATGCA CACGGTATCC TCTGCGGGTT TAGGCATTAA AACCTCTCCT	6780
ACACAAAATC TAAGCCTAGA TGCTTTGTT GCTCGTCGCT TTGCAAATGC CAATAGTGAC	6840
AATTGAAATG GCAACAAAAA ACGCACAAAGC TCACCTACAA CCTTCTGGGG GAGATTAACA	6900
TTCAGTTCT AACCGTGAAC TTTAATCAAC TGGTAAGCGT TCCGCCTACC AGTTTATAAC	6960
TATATGCTTT ACCCGCCAAT TTACAGTCTA TAGGCAACCC TGTTTTTACC CTTATATATC	7020
AAATAAACAA GCTAAGCTGA GCTAAGCAA CCAAGCAAAC TCAAGCAAGC CAAGTAATAC	7080
TAAAAAAACA ATTTATATGA TAAACTAAAG TATACTCCAT GCCATGGCGA TACAAGGGAT	7140
TTAATAATAT GACAAAAGAA AATTGCAAAC ACGCTCCTCA AGATGCGACC GCTTTACTTG	7200
CGGAATTAAG CAACAATCAA ACTCCCTGC GAATATTTAA ACAACCACGC AAGCCCAGCC	7260
TATTACGCTT GGAACAACAT ATCGCAAAAAA AAGATTATGA GTTGTGTTGT CGTGAATTAA	7320
TGGTGATTCT GGAAAAAAATG GACGCTAATT TTGGAGGGCGT TCACGATATT GAATTTGACG	7380
CACCCGCTCA GCTGGCATAT CTACCCGAAA AATTACTAAT TTATTTGCC ACTCGTCTCG	7440
CTAATGCAAT TACAACACTC TTTTCCGACC CCGAATTGGC AATTCTGAA GAAGGGCGT	7500
TAAAGATGAT TAGCCTGCAA CGCTGGTTGA CGCTGATTT TGCTCTTCC CCCTACGTTA	7560
ACGCAGACCA TATTCTCAAT AAATATAATA TCAACCCAGA TTCCGAAGGT GGCTTTCATT	7620
TAGCAACAGA CAACTCTCT ATTGCTAAAT TCTGTATTT TTACTTACCC GAATCCAATG	7680
TCAATATGAG TTTAGATGCG TTATGGGCAG GGAATCAACA ACTTTGTGCT TCATTGTGTT	7740
TTGCGTTGCA GTCTTCACGT TTATTTGGTA CCGCATCTGC GTTTCATAAA AGAGCGGTGG	7800
TTTACAGTG GTTTCCTAAA AAACTCGCCG AAATTGCTAA TTTAGATGAA TTGCGCTGCAA	7860
ATATCCTTCA TGATGTATAT ATGCACTGCA GTTATGATT AGCAAAAAAC AAGCACGATG	7920
TTAAGCGTCC ATTAAACGAA CTTGTCCGCA AGCATATCCT CACGCAAGGA TGGCAAGACC	7980
GCTACCTTTA CACCTTAGGT AAAAAGGACG GCAAACCTGT GATGATGGTA CTGCTTGAAC	8040
ATTTTAATTC GGGACATTG ATTTATCGTA CACATTCAAC TTCAATGATT GCTGCTCGAG	8100
AAAAATTCTA TTTAGTCGGC TTAGGCCATG AGGGCGTTGA TAAAATAGGT CGAGAAGTGT	8160
TTGACGAGTT CTTTGAAATC AGTAGCAATA ATATAATGGA GAGACTGTTT TTTATCCGTA	8220

AACAGTGCAGA AACTTTCCAA CCCGCAGTGT TCTATATGCC AAGCATTGGC ATGGATATTA	8280
CCACGATTT TGTGAGCAAC ACTCGGCTTG CCCCTATTCA AGCTGTAGCC CTGGGTCATC	8340
CTGCCACTAC GCATTCTGAA TTTATTGATT ATGTCATCGT AGAAGATGAT TATGTGGCA	8400
GTGAAGATTG TTTCAGCGAA ACCCTTTAC GCTTACCCAA AGATGCCCTA CCTTATGTAC	8460
CTTCTGCACT CGCCCCACAA AAAGTGGATT ATGTAACAG GGAAAACCCCT GAAGTAGTCA	8520
ATATCGGTAT TGCCGCTACC ACAATGAAAT TAAACCCCTGA ATTTTGCTA ACATTGCAAG	8580
AAATCAGAGA TAAAGCTAAA GTCAAAATAC ATTTTCATT CGCACTTGGA CAATCAACAG	8640
GCTTGACACA CCCTTATGTC AAATGGTTA TCGAAAGCTA TTTAGGTGAC GATGCCACTG	8700
CACATCCCCA CGCACCTTAT CACGATTATC TGGCAATATT GCGTGATTGC GATATGCTAC	8760
TAAATCCGTT TCCTTTCGGT AATACTAACG GCATAATTGA TATGGTTACA TTAGGTTAG	8820
TTGGTGTATG CAAAACGGGG GATGAAGTAC ATGAACATAT TGATGAAGGT CTGTTAAC	8880
GCTTAGGACT ACCAGAATGG CTGATAGCCG ACACACGAGA AACATATATT GAATGTGCTT	8940
TGCGTCTAGC AGAAAACCAT CAAGAACGCC TTGAACCTCG TCGTTACATC ATAGAAAACA	9000
ACGGCTTACA AAAGCTTTT ACAGGCGACC CTCGTCCATT GGGCAAAATA CTGCTTAAGA	9060
AAACAAATGA ATGGAAGCGG AAGCACTTGA GTAAAAAATA ACGGTTTTT AAAGTAAAAG	9120
TGCGTTAAT TTTCAAAGCG TTTTAAAAAC CTCTCAAAAA TCAACCGCAC TTTTATCTTT	9180
ATAACGATCC CGCACGCTGA CAGTTTATCA GCCTCCCGCC ATAAAACCTCC GCCTTCATG	9240
GCGGAGATTT TAGCCAAAAC TGGCAGAAAT TAAAGGCTAA AATCACCAAA TTGCACCACA	9300
AAATCACCAA TACCCACAAA AAA	9323

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4794 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGAACAAGA TATATCGTCT CAAATTCAAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT	60
GAATTGACAC GGGGTTGTGA CCATTCCACA GAAAAAGGCA GTGAAAAACC TGTCGTACG	120
AAAGTACGCC ACTTGGCGTT AAAGCCACTT TCCGCTATAT TGCTATCTTT GGGCATGGCA	180
TCCATTCCGC AATCTGTTTT AGCGAGCGGT TTACAGGGAA TGAGCGTCGT ACACGGTACA	240
GCAACCATGC AAGTAGACGG CAATAAAACC ACTATCCGTA ATAGCGTCAA TGCTATCATC	300
AATTGGAAAC AATTAAACAT TGACCCAAAT GAAATGGTGC AGTTTTACA AGAAAGCAGC	360
AACTCTGCCG TTTTCAACCG TGTTACATCT GACCAAATCT CCCAATTAAA AGGGATTTA	420

GATTCTAACG GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA	480
ATTATTAACA CTAATGGCTT TACTGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG	540
GCGCGTAATT TCACCCCTGGA GCAAACCAAG GATAAAGCAC TCGCTGAAAT CGTGAATCAC	600
GGTTTAATTA CCGTTGGTAA AGACGGTAGC GTAAACCTTA TTGGTGGCAA AGTGAACAAAC	660
GAGGGCGTGA TTAGCGTAAA TGGCGGTAGT ATTTCTTAC TTGCAGGGCA AAAATCACC	720
ATCAGCGATA TAATAAATCC AACCATCACT TACAGCATTG CTGCACCTGA AAACGAAGCG	780
ATCAATCTGG GCGATATTT TGCCAAAGGT GGTAAACATTA ATGTCCGCGC TGCCACTATT	840
CGCAATAAAG GTAAACTTTC TGCCGACTCT GTAAAGCAAAG ATAAAAGTGG TAACATTGTT	900
CTCTCTGCCA AAGAAGGTGA AGCGGAAATT GGCGGTGTA TTTCCGCTCA AAATCAGCAA	960
GCCAAAGGTG GTAAAGTGT GATTACAGGC GATAAAGTTA CATTGAAAAC GGGTGCAGTT	1020
ATCGACCTTT CGGGTAAAGA AGGGGGAGAA ACTTATCTTG GCGGTGACGA GCGTGGCGAA	1080
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCACTTTAG AAAAAGGCTC AACAAATTAAT	1140
GTGTCAGGTA AAGAAAAAGG TGGCGCGCT ATTGTATGGG GCGATATTGC GTTAATTGAC	1200
GGCAATATTA ATGCCCAAGG TAAAGATATC GCTAAAACGT GTGGTTTGT GGAGACGTCG	1260
GGGCATTACT TATCCATTGA TGATAACGCA ATTGTTAAAA CAAAAGAATG GCTACTAGAC	1320
CCAGAGAATG TGACTATTGA AGCTCCTTCC GCTTCTCGCG TCGAGCTGGG TGCCGATAGG	1380
AATTCCCACG CGGCAGAGGT GATAAAAGTG ACCCTAAAAA AAAATAACAC CTCCTTGACA	1440
ACACTAACCA ATACAACCAT TTCAAATCTT CTGAAAAGTG CCCACGTGGT GAACATAACG	1500
GCAAGGAGAA AACTTACCGT TAATAGCTCT ATCAGTATAG AAAGAGGCTC CCACCTTAATT	1560
CTCCACAGTG AAGGTCAAGGG CGGTCAAGGT GTTCAGATTG ATAAAGATAT TACTTCTGAA	1620
GGCGGAAATT TAACCATTAA TTCTGGCGGA TGGGTTGATG TTCATAAAAA TATTACGCTT	1680
GGTAGCGGCT TTTTAAACAT CACAACAAA GAAGGAGATA TCGCCTTCGA AGACAAGTCT	1740
GGACGGAACA ACCTAACCAT TACAGCCAA GGGACCATCA CCTCAGGTAA TAGTAACGGC	1800
TTTAGATTAA ACAACGTCTC TCTAACAGC CTTGGCGGAA AGCTGAGCTT TACTGACAGC	1860
AGAGAGGACA GAGGTAGAAG AACTAAGGGT AATATCTCAA ACAAAATTGTA CGGAACGTTA	1920
AACATTCCG GAACTGTAGA TATCTCAATG AAAGCACCCCA AAGTCAGCTG GTTTTACAGA	1980
GACAAAGGAC GCACCTACTG GAACGTAACC ACTTTAAATG TTACCTCGGG TAGTAAATT	2040
AACCTCTCCA TTGACAGCAC AGGAAGTGGC TCAACAGGTC CAAGCATAACG CAATGCAGAA	2100
TTAAATGGCA TAACATTAA TAAAGCCACT TTTAATATCG CACAAGGCTC AACAGCTAAC	2160
TTTAGCATCA AGGCATCAAT AATGCCCTTT AAGAGTAACG CTAACATACGC ATTATTTAAT	2220
GAAGATATTG CAGTCTCAGG GGGGGTAGC CTTAATTCA AACTAACGC CTCATCTAGC	2280
AACATACAAA CCCCTGGCGT AATTAAAAA TCTCAAAACT TTAATGTCTC AGGAGGGTCA	2340
ACTTTAAATC TCAAGGCTGA AGGTTCAACA GAAACCGCTT TTTCAATAGA AAATGATTAA	2400
AACTTAAACG CCACCGGTGG CAATATAACA ATCAGACAAG TCGAGGGTAC CGATTCACGC	2460

GTCAACAAAG	GTGTCGCAGC	CAAAAAAAAC	ATAACTTTA	AAGGGGTAA	TATCACCTTC	2520
GGCTCTAAA	AAGCCACAAAC	AGAAATCAA	GGCAATGTTA	CCATCAATAA	AAACACTAAC	2580
GCTACTCTT	GTGGTGCAGA	TTTGCCGAA	AACAAATCGC	CTTTAAATAT	AGCAGGAAAT	2640
GTTATTAATA	ATGGCAACCT	TACCACTGCC	GGCTCCATTA	TCAATATAGC	CGGAAATCTT	2700
ACTGTTCAA	AAGGCAGCTAA	CCTTCAAGCT	ATAACAAATT	ACACTTTAA	TGTAGCCGGC	2760
TCATTTGACA	ACAATGGCGC	TTCAAACATT	TCCATTGCCA	GAGGAGGGC	TAAATTTAAA	2820
GATATCAATA	ACACCAGTAG	CTTAAATATT	ACCACCAA	CTGATAACCAC	TTACCGCACC	2880
ATTATAAAAG	GCAATATATC	CAACAAATCA	GGTGATTG	ATATTATTGA	TAAAAAAAGC	2940
GACGCTGAAA	TCCAAATTGG	CGGCAATATC	TCACAAAAAG	AAGGCAATCT	CACAATTCT	3000
TCTGATAAAAG	TAAATATTAC	CAATCAGATA	ACAATCAAAG	CAGGCGTTGA	AGGGGGCGT	3060
TCTGATTCAA	GTGAGGCAGA	AAATGCTAAC	CTAACTATT	AAACCAAAGA	GTAAATTTG	3120
GCAGGAGACC	TAAATATTTC	AGGCTTTAAT	AAAGCAGAAA	TTACAGCTAA	AAATGGCAGT	3180
GATTAACTA	TTGGCAATGC	TAGCGGTGGT	AATGCTGATG	CTAAAAAGT	GACTTTGAC	3240
AAGGTTAAAG	ATTCAAAAAT	CTCGACTGAC	GGTCACAATG	TAACACTAA	TAGCGAAGTG	3300
AAAACGTCTA	ATGGTAGTAG	CAATGCTGGT	AATGATAACA	GCACCGTTT	AACCATTCC	3360
GCAAAAGATG	TAACGGTAAA	CAATAACGTT	ACCTCCCACA	AGACAATAAA	TATCTCTGCC	3420
GCAGCAGGAA	ATGTAACAAAC	CAAAGAAGGC	ACAACATATCA	ATGCAACCAC	AGGCAGCGTG	3480
GAAGTAACTG	CTCAAAATGG	TACAATTAAA	GGCAACATTA	CCTCGAAAA	TGTAACAGTG	3540
ACAGCAACAG	AAAATCTTGT	TACCACAGAG	AATGCTGTCA	TTAATGCAAC	CAGCGGCACA	3600
GTAAACATTA	GTACAAAAAC	AGGGATATT	AAAGGTGGAA	TTGAATCAAC	TTCCGGTAAT	3660
GTAAATATTA	CAGCGAGCGG	CAATACACTT	AAGGTAAGTA	ATATCACTGG	TCAAGATGTA	3720
ACAGTAACAG	CGGATGCAGG	AGCCTTGACA	ACTACAGCAG	GCTCAACCAC	TAGTGCAGACA	3780
ACAGGCAATG	CAAATATTAC	AACCAAAACA	GGTGATATCA	ACGGTAAAGT	TGAATCCAGC	3840
TCCGGCTCTG	TAACACTTGT	TGCAACTGGA	GCAACTCTTG	CTGTAGGTAA	TATTCAGGT	3900
AACACTGTTA	CTATTACTGC	GGATAGCGGT	AAATTAACCT	CCACAGTAGG	TTCTACAATT	3960
AATGGGACTA	ATAGTGTAAAC	CACCTCAAGC	CAATCAGGCG	ATATTGAAGG	TACAATTCT	4020
GGTAATACAG	TAAATGTTAC	AGCAAGCACT	GGTGATTAA	CTATTGAAA	TAGTGCAAAA	4080
GTTGAAGCGA	AAAATGGAGC	TGCAACCTTA	ACTGCTGAAT	CAGGCAAATT	AACCACCCAA	4140
ACAGGCTCTA	GCATTACCTC	AAGCAATGGT	CAGACAACTC	TTACAGCCAA	GGATAGCGAGT	4200
ATCGCAGGAA	ACATTAATGC	TGCTAATGTG	ACGTTAAATA	CCACAGGCAC	TTTAACTACT	4260
ACAGGGGATT	CAAAGATTAA	CGCAACCAGT	GGTACCTAA	CAATCAATGC	AAAAGATGCC	4320
AAATTAGATG	GTGCTGCATC	AGGTGACCGC	ACAGTAGTAA	ATGCAACTAA	CGCAAGTGGC	4380
TCTGGTAACG	TGACTGCGAA	AACCTCAAGC	AGCGTGAATA	TCACCGGGG	TTTAAACACA	4440
ATAAATGGGT	TAAATATCAT	TTCGGAAAAT	GGTAGAAACA	CTGTGCCTT	AAGAGGCAAG	4500

GAAATTGATG TGAAATATAT CCAACCAGGT GTAGCAAGCG TAGAAGAGGT AATTGAAGCG	4560
AAACCGTCC TTGAGAAGGT AAAAGATTTA TCTGATGAAG AAAGAGAAAC ACTAGCCAAA	4620
CTTGGTGTAA GTGCTGTACG TTTCGTTGAG CCAAATAATG CCATTACGGT TAATACACAA	4680
AACGAGTTA CAACCAAACC ATCAAGTCAA GTGACAATTCTGAAGGTAA GGCGTGTTC	4740
TCAAGTGGTA ATGGCGCACG AGTATGTACC AATGTTGCTG ACGATGGACA GCAG	4794

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4803 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGAACAAGA TATATCGTCT CAAATTCAAGC AAACGCCCTGA ATGCTTTGGT TGCTGTGTCT	60
GAATTGACAC GGGGTTGTGA CCATTCCACA GAAAAAGGCA GTGAAAACC TGTCGTACG	120
AAAGTACGCC ACTTGGCGTT AAAGCCACTT TCCGCTATAT TGCTATCTT GGGCATGGCA	180
TCCATTCCGC AATCTGTTT AGCGAGCGGT TTACAGGGAA TGAGCGTCGT ACACGGTACA	240
GCAACCATGC AAGTAGACGG CAATAAAACC ACTATCCGTA ATAGCGTCAA TGCTATCATC	300
AATTGGAAAC AATTAAACAT TGACCAAAAT GAAATGGTGC AGTTTTACA AGAAAGCAGC	360
AACTCTGCCG TTTTCAACCG TGTTACATCT GACCAAATCT CCCAATTAAA AGGGATTTA	420
GATTCTAACG GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA	480
ATTATTAACA CTAATGGCTT TACTGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG	540
GCGCGTAATT TCACCCCTGA GCAAACCAAG GATAAAGCAC TCGCTGAAAT CGTGAATCAC	600
GGTTTAATT CCAGTTGGTAA AGACGGTAGC GTAAACCTTA TTGGTGGCAA AGTGAACAC	660
GAGGGCGTGA TTAGCGTAAA TGGCGGTAGT ATTCTTTAC TTGCAGGGCA AAAAATCACC	720
ATCAGCGATA TAATAAATCC AACCCTCACT TACAGCATTG CTGCACCTGA AAACGAAGCG	780
ATCAATCTGG GCGATATTTT TGCCAAAGGT GGTAACATTA ATGTCCCGC TGCCACTATT	840
CGCAATAAAG GTAAACTTTC TGCCGACTCT GTAAGCAAAG ATAAAAGTGG TAACATTGTT	900
CTCTCTGCCA AAGAAGGTGA AGCGGAAATT GGCGGTGTAA TTTCCGCTCA AAATCAGCAA	960
GCCAAAGGTG GTAAGTTGAT GATTACAGGT GATAAAGTCAT CATTAAAAAC AGGTGCAGTT	1020
ATCGACCTTT CAGGTAAAGA AGGGGGAGAG ACTTATCTTG GCGGTGATGA GCGTGGCGAA	1080
GGTAAAAATG GTATTCAATT AGCGAAGAAA ACCTCTTAG AAAAAGGCTC GACAATTAAT	1140
GTATCAGGCA AAGAAAAAGG CGGGCGCGCT ATTGTATGGG GCGATATTGC ATTAATTAAT	1200
GGTAACATTA ATGCTCAAGG TAGCGATATT GCTAAAATG GCGGCTTGT GGAAACATCA	1260

GGACATGACT TATCCATTGG TGATGATGTG ATTGTTGACG CTAAGAGTG GTTATTAGAC	1320
CCAGATGATG TGTCCATTGA AACTCTTACA TCTGGACGCA ATAATACCGG CGAAAACCAA	1380
GGATATAACAA CAGGAGATGG GACTAAAGAG TCACCTAAAG GTAATAGTAT TTCTAACCT	1440
ACATTAACAA ACTCAACTCT TGAGCAAATC CTAAGAAGAG GTTCTTATGT TAATATCACT	1500
GCTAATAATA GAATTTATGT TAATAGCTCC ATCAACTTAT CTAATGGCAG TTTAACACTT	1560
CACACTAAAC GAGATGGAGT TAAAATTAAC GGTGATATTA CCTCAAACGA AAATGGTAAT	1620
TTAACCATTA AAGCAGGCTC TTGGGTTGAT GTTCATAAAA ACATCACGCT TGGTACGGGT	1680
TTTTGAATA TTGTCGCTGG GGATTCTGTA GCTTTGAGA GAGAGGGCGA TAAAGCACGT	1740
AACGCAACAG ATGCTCAAAT TACCGCACAA GGGACGATAA CCGTCAATAA AGATGATAAA	1800
CAATTTAGAT TCAATAATGT ATCTATTAAC GGGACGGGCA AGGGTTAAA GTTTATTGCA	1860
AATCAAAATA ATTCACACTCA TAAATTTGAT GGCAGAAATTAA ACATATCTGG AATAGTAACA	1920
ATTAACCAA CCACGAAAAA AGATGTTAAA TACTGGAATG CATCAAAAGA CTCTTACTGG	1980
AATGTTCTT CTCTTACTTT GAATACGGTG CAAAAATTAA CCTTTATAAA ATTGTTGAT	2040
AGCGGCTCAA ATTCCCAAGA TTTGAGGTCA TCACGTAGAA GTTTGCGAGG CGTACATT	2100
AACGGCATCG GAGGCAAAAC AAACCTCAAC ATCGGAGCTA ACGCAAAAGC CTTATTAA	2160
TTAAAACCAA ACGCCGCTAC AGACCCAAA AAAGAATTAC CTATTACTTT TAACGCCAAC	2220
ATTACAGCTA CCGGTAACAG TGATAGCTCT GTGATGTTG ACATACACGC CAATCTTACC	2280
TCTAGAGCTG CCGGCATAAA CATGGATTCA ATTAACATTA CCGGCGGGCT TGACTTTCC	2340
ATAACATCCC ATAATCGCAA TAGTAATGCT TTTGAAATCA AAAAGACTT AACTATAAT	2400
GCAACTGGCT CGAATTTAG TCTTAAGCAA ACGAAAGATT CTTTTATAAA TGAATACAGC	2460
AAACACGCCA TTAACTCAAG TCATAATCTA ACCATTCTG GCGCAATGT CACTCTAGGT	2520
GGGGAAAATT CAAGCAGTAG CATTACGGC AATATCAATA TCACCAATAA AGCAAATGTT	2580
ACATTACAAG CTGACACCAG CAACAGCAAC ACAGGCTTGA AGAAAAGAAC TCTAATCTT	2640
GGCAATATAT CTGTTGAGGG GAATTAAAGC CTAACTGGTG CAAATGCAA CATTGTCGGC	2700
AATCTTTCTA TTGCAGAAGA TTCCACATTT AAAGGAGAAG CCAGTGACAA CCTAAACATC	2760
ACCGGCACCT TTACCAACAA CGGTACCGCC AACATTAATA TAAAACAAG AGTGGTAAA	2820
CTCCAAGGCG ATATTATCAA TAAAGGTGGT TAAATATCA CTACTAACGC CTCAGGCACT	2880
CAAAAAACCA TTATTAACGG AAATATAACT AACGAAAAG GCGACTTAA CATCAAGAAT	2940
ATTAAAGCCG ACGCCGAAAT CCAAATTGGC GCGAATATCT CACAAAAGA AGGCAATCTC	3000
ACAATTCTT CTGATAAAAGT AAATATTACC AATCAGATAA CAATCAAAGC AGGCAGTTGAA	3060
GGGGGGCGTT CTGATTCAAG TGAGGCAGAA AATGCTAACCTAACTATTCA AACCAAAGAG	3120
TTAAAATTGG CAGGAGACCT AAATATTCA GGCTTTAATA AAGCAGAAAT TACAGCTAAA	3180
AATGGCAGTG ATTTAACTAT TGGCAATGCT AGCGGTGGTA ATGCTGATGC TAAAAAAAGTG	3240
ACTTTGACA AGGTTAAAGA TTCAAAATC TCGACTGACG GTCACAATGT AACACTAAAT	3300

AGCGAAGTGA	AAACGTCTAA	TGGTAGTAGC	AATGCGGGTA	ATGATAACAG	CACCGGTTA	3360
ACCATTTCCG	CAAAAGATGT	AACGGTAAAC	AATAAAGTTA	CCTCCCACAA	GACAATAAAT	3420
ATCTCTGCCG	CAGCAGGAAA	TGTAACAACC	AAAGAAGGCA	CAACTATCAA	TGCAACCACA	3480
GGCAGCGTGG	AAGTAACACTGC	TCAAAATGGT	ACAAATAAAG	GCAACATTAC	CTCGCAAAAT	3540
GTAACAGTGA	CAGCAACAGA	AAATCTTGT	ACCACAGAGA	ATGCTGTCA	TAATGCAACC	3600
AGCGGCACAG	TAAACATTAG	TACAAAACA	GGGGATATTA	AAGGTGGAAT	TGAATCAACT	3660
TCCGGTAATG	TAAATATTAC	AGCGAGCGGC	AATACACTTA	AGGTAAGTAA	TATCACTGGT	3720
CAAGATGTAA	CAGTAACAGC	GGATGCAGGA	GCCTTGACAA	CTACAGCAGG	CTCAACCATT	3780
AGTGCACAA	CAGGCAATGC	AAATATTACA	ACCAAAACAG	GTGATATCAA	CGGTAAAGTT	3840
GAATCCAGCT	CCGGCTCTGT	AACACTTGT	GCAACTGGAG	CAACTCTTC	TGTAGGTAAT	3900
ATTCAGGTA	ACACTGTTAC	TATTACTGCG	GATAGCGGT	AATTAAACCTC	CACAGTAGGT	3960
TCTACAATT	ATGGGACTAA	TAGTGTAA	ACCTCAAGCC	AATCAGGCGA	TATTGAAGGT	4020
ACAATTCTG	GTAATACAGT	AAATGTTACA	GCAAGCACTG	GTGATTTAAC	TATTGAAAT	4080
AGTGCAAAAG	TTGAAGCGAA	AAATGGAGCT	GCAACCTTAA	CTGCTGAATC	AGGCAAATTA	4140
ACCACCCAAA	CAGGCTCTAG	CATTACCTCA	AGCAATGGTC	AGACAACTCT	TACAGCCAAG	4200
GATAGCAGTA	TCGCAGGAAA	CATTAATGCT	GCTAATGTGA	CGTTAAATAC	CACAGGCACT	4260
TTAACTACTA	CAGGGGATTC	AAAGATTAAC	GCAACCAGTG	GTACCTTAAC	AATCAATGCA	4320
AAAGATGCCA	AATTAGATGG	TGCTGCATCA	GGTGACCGCA	CAGTAGTAAA	TGCAACTAAC	4380
GCAAGTGGCT	CTGGTAACGT	GACTGCGAAA	ACCTCAAGCA	GCGTGAATAT	CACCGGGGAT	4440
TTAACACAA	AAAATGGGTT	AAATATCATT	TCGGAAAATG	GTAGAAACAC	TGTGCGCTTA	4500
AGAGGCAAGG	AAATTGATGT	GAAATATATC	CAACCAGGTG	TAGCAAGCGT	AGAAGAGGTA	4560
ATTGAAGCGA	AACCGTCCCT	TGAGAAGGTA	AAAGATTTAT	CTGATGAAGA	AAGAGAAACA	4620
CTAGCCAAAC	TTGGTGTAAAG	TGCTGTACGT	TTCTGTTGAGC	CAAATAATGC	CATTACGGTT	4680
AATACACAAA	ACGAGTTTAC	AACCAAACCA	TCAAGTCAG	TGACAATTTC	TGAAGGTAAG	4740
GCGTGTTC	CAAGTGGTAA	TGGCGCACGA	GTATGTACCA	ATGTTGCTGA	CGATGGACAG	4800
CAG						4803

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1599 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300

Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320

Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Lys Asp Ile Ala Lys Thr Gly Phe
 405 410 415
 Val Glu Thr Ser Gly His Tyr Leu Ser Ile Asp Asp Asn Ala Ile Val
 420 425 430
 Lys Thr Lys Glu Trp Leu Leu Asp Pro Glu Asn Val Thr Ile Glu Ala
 435 440 445
 Pro Ser Ala Ser Arg Val Glu Leu Gly Ala Asp Arg Asn Ser His Ser
 450 455 460
 Ala Glu Val Ile Lys Val Thr Leu Lys Asn Asn Thr Ser Leu Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Ile Ser Asn Leu Leu Lys Ser Ala His Val
 485 490 495
 Val Asn Ile Thr Ala Arg Arg Lys Leu Thr Val Asn Ser Ser Ile Ser
 500 505 510
 Ile Glu Arg Gly Ser His Leu Ile Leu His Ser Glu Gly Gln Gly Gly
 515 520 525
 Gln Gly Val Gln Ile Asp Lys Asp Ile Thr Ser Glu Gly Asn Leu
 530 535 540
 Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr Leu
 545 550 555 560
 Gly Ser Gly Phe Leu Asn Ile Thr Thr Lys Glu Gly Asp Ile Ala Phe
 565 570 575
 Glu Asp Lys Ser Gly Arg Asn Asn Leu Thr Ile Thr Ala Gln Gly Thr
 580 585 590
 Ile Thr Ser Gly Asn Ser Asn Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Ser Leu Gly Gly Lys Leu Ser Phe Thr Asp Ser Arg Glu Asp Arg
 610 615 620
 Gly Arg Arg Thr Lys Gly Asn Ile Ser Asn Lys Phe Asp Gly Thr Leu
 625 630 635 640
 Asn Ile Ser Gly Thr Val Asp Ile Ser Met Lys Ala Pro Lys Val Ser
 645 650 655
 Trp Phe Tyr Arg Asp Lys Gly Arg Thr Tyr Trp Asn Val Thr Thr Leu
 660 665 670
 Asn Val Thr Ser Gly Ser Lys Phe Asn Leu Ser Ile Asp Ser Thr Gly
 675 680 685

Ser Gly Ser Thr Gly Pro Ser Ile Arg Asn Ala Glu Leu Asn Gly Ile
 690 695 700
 Thr Phe Asn Lys Ala Thr Phe Asn Ile Ala Gln Gly Ser Thr Ala Asn
 705 710 715 720
 Phe Ser Ile Lys Ala Ser Ile Met Pro Phe Lys Ser Asn Ala Asn Tyr
 725 730 735
 Ala Leu Phe Asn Glu Asp Ile Ser Val Ser Gly Gly Ser Val Asn
 740 745 750
 Phe Lys Leu Asn Ala Ser Ser Asn Ile Gln Thr Pro Gly Val Ile
 755 760 765
 Ile Lys Ser Gln Asn Phe Asn Val Ser Gly Gly Ser Thr Leu Asn Leu
 770 775 780
 Lys Ala Glu Gly Ser Thr Glu Thr Ala Phe Ser Ile Glu Asn Asp Leu
 785 790 795 800
 Asn Leu Asn Ala Thr Gly Gly Asn Ile Thr Ile Arg Gln Val Glu Gly
 805 810 815
 Thr Asp Ser Arg Val Asn Lys Gly Val Ala Ala Lys Lys Asn Ile Thr
 820 825 830
 Phe Lys Gly Gly Asn Ile Thr Phe Gly Ser Gln Lys Ala Thr Thr Glu
 835 840 845
 Ile Lys Gly Asn Val Thr Ile Asn Lys Asn Thr Asn Ala Thr Leu Arg
 850 855 860
 Gly Ala Asn Phe Ala Glu Asn Lys Ser Pro Leu Asn Ile Ala Gly Asn
 865 870 875 880
 Val Ile Asn Asn Gly Asn Leu Thr Thr Ala Gly Ser Ile Ile Asn Ile
 885 890 895
 Ala Gly Asn Leu Thr Val Ser Lys Gly Ala Asn Leu Gln Ala Ile Thr
 900 905 910
 Asn Tyr Thr Phe Asn Val Ala Gly Ser Phe Asp Asn Asn Gly Ala Ser
 915 920 925
 Asn Ile Ser Ile Ala Arg Gly Gly Ala Lys Phe Lys Asp Ile Asn Asn
 930 935 940
 Thr Ser Ser Leu Asn Ile Thr Thr Asn Ser Asp Thr Thr Tyr Arg Thr
 945 950 955 960
 Ile Ile Lys Gly Asn Ile Ser Asn Lys Ser Gly Asp Leu Asn Ile Ile
 965 970 975
 Asp Lys Lys Ser Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln
 980 985 990
 Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr Asn
 995 1000 1005
 Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser Ser
 1010 1015 1020
 Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys Leu
 1025 1030 1035 1040

Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala
 1045 1050 1055
 Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn Ala
 1060 1065 1070
 Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser Asn
 1090 1095 1100
 Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile Ser
 1105 1110 1115 1120
 Ala Lys Asp Val Thr Val Asn Asn Val Thr Ser His Lys Thr Ile
 1125 1130 1135
 Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr Thr
 1140 1145 1150
 Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly Thr
 1155 1160 1165
 Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr Glu
 1170 1175 1180
 Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly Thr
 1185 1190 1195 1200
 Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu Ser
 1205 1210 1215
 Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys Val
 1220 1225 1230
 Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly Ala
 1235 1240 1245
 Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn Ala
 1250 1255 1260
 Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser Ser
 1265 1270 1275 1280
 Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val Gly
 1285 1290 1295
 Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys Leu
 1300 1305 1310
 Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr Thr
 1315 1320 1325
 Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr Val
 1330 1335 1340
 Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala Lys
 1345 1350 1355 1360
 Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly Lys
 1365 1370 1375
 Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln Thr
 1380 1385 1390

Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala Ala
 1395 1400 1405
 Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Gly Asp Ser
 1410 1415 1420
 Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp Ala
 1425 1430 1435 1440
 Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala Thr
 1445 1450 1455
 Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Val
 1460 1465 1470
 Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile Ser
 1475 1480 1485
 Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp Val
 1490 1495 1500
 Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu Ala
 1505 1510 1515 1520
 Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg Glu
 1525 1530 1535
 Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro Asn
 1540 1545 1550
 Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro Ser
 1555 1560 1565
 Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly Asn
 1570 1575 1580
 Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1600 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15
 Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30
 Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45
 Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95
 Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110
 Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
 115 120 125
 Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Asp Ile Ala Lys Thr Gly Gly Phe
 405 410 415
 Val Glu Thr Ser Gly His Asp Leu Ser Ile Gly Asp Asp Val Ile Val
 420 425 430

Asp Ala Lys Glu Trp Leu Leu Asp Pro Asp Asp Val Ser Ile Glu Thr
 435 440 445
 Leu Thr Ser Gly Arg Asn Asn Thr Gly Glu Asn Gln Gly Tyr Thr Thr
 450 455 460
 Gly Asp Gly Thr Lys Glu Ser Pro Lys Gly Asn Ser Ile Ser Lys Pro
 465 470 475 480
 Thr Leu Thr Asn Ser Thr Leu Glu Gln Ile Leu Arg Arg Gly Ser Tyr
 485 490 495
 Val Asn Ile Thr Ala Asn Asn Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu His Thr Lys Arg Asp Gly Val Lys
 515 520 525
 Ile Asn Gly Asp Ile Thr Ser Asn Glu Asn Gly Asn Leu Thr Ile Lys
 530 535 540
 Ala Gly Ser Trp Val Asp Val His Lys Asn Ile Thr Leu Gly Thr Gly
 545 550 555 560
 Phe Leu Asn Ile Val Ala Gly Asp Ser Val Ala Phe Glu Arg Glu Gly
 565 570 575
 Asp Lys Ala Arg Asn Ala Thr Asp Ala Gln Ile Thr Ala Gln Gly Thr
 580 585 590
 Ile Thr Val Asn Lys Asp Asp Lys Gln Phe Arg Phe Asn Asn Val Ser
 595 600 605
 Leu Asn Gly Thr Gly Lys Gly Leu Lys Phe Ile Ala Asn Gln Asn Asn
 610 615 620
 Phe Thr His Lys Phe Asp Gly Glu Ile Asn Ile Ser Gly Ile Val Thr
 625 630 635 640
 Ile Asn Gln Thr Thr Lys Lys Asp Val Lys Tyr Trp Asn Ala Ser Lys.
 645 650 655
 Asp Ser Tyr Trp Asn Val Ser Ser Leu Thr Leu Asn Thr Val Gln Lys
 660 665 670
 Phe Thr Phe Ile Lys Phe Val Asp Ser Gly Ser Asn Gly Gln Asp Leu
 675 680 685
 Arg Ser Ser Arg Arg Ser Phe Ala Gly Val His Phe Asn Gly Ile Gly
 690 695 700
 Gly Lys Thr Asn Phe Asn Ile Gly Ala Asn Ala Lys Ala Leu Phe Lys
 705 710 715 720
 Leu Lys Pro Asn Ala Ala Thr Asp Pro Lys Lys Glu Leu Pro Ile Thr
 725 730 735
 Phe Asn Ala Asn Ile Thr Ala Thr Gly Asn Ser Asp Ser Ser Val Met
 740 745 750
 Phe Asp Ile His Ala Asn Leu Thr Ser Arg Ala Ala Gly Ile Asn Met
 755 760 765
 Asp Ser Ile Asn Ile Thr Gly Gly Leu Asp Phe Ser Ile Thr Ser His
 770 775 780

100

Asn Arg Asn Ser Asn Ala Phe Glu Ile Lys Lys Asp Leu Thr Ile Asn
 785 790 795 800
 Ala Thr Gly Ser Asn Phe Ser Leu Lys Gln Thr Lys Asp Ser Phe Tyr
 805 810 815
 Asn Glu Tyr Ser Lys His Ala Ile Asn Ser Ser His Asn Leu Thr Ile
 820 825 830
 Leu Gly Gly Asn Val Thr Leu Gly Gly Glu Asn Ser Ser Ser Ser Ile
 835 840 845
 Thr Gly Asn Ile Asn Ile Thr Asn Lys Ala Asn Val Thr Leu Gln Ala
 850 855 860
 Asp Thr Ser Asn Ser Asn Thr Gly Leu Lys Lys Arg Thr Leu Thr Leu
 865 870 875 880
 Gly Asn Ile Ser Val Glu Gly Asn Leu Ser Leu Thr Gly Ala Asn Ala
 885 890 895
 Asn Ile Val Gly Asn Leu Ser Ile Ala Glu Asp Ser Thr Phe Lys Gly
 900 905 910
 Glu Ala Ser Asp Asn Leu Asn Ile Thr Gly Thr Phe Thr Asn Asn Gly
 915 920 925
 Thr Ala Asn Ile Asn Ile Lys Gly Val Val Lys Leu Gly Asp Ile Asn
 930 935 940
 Asn Lys Gly Gly Leu Asn Ile Thr Thr Asn Ala Ser Gly Thr Gln Lys
 945 950 955 960
 Thr Ile Ile Asn Gly Asn Ile Thr Asn Glu Lys Gly Asp Leu Asn Ile
 965 970 975
 Lys Asn Ile Lys Ala Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr
 995 1000 1005
 Asn Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser
 1010 1015 1020
 Ser Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn
 1060 1065 1070
 Ala Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile
 1075 1080 1085
 Ser Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser
 1090 1095 1100
 Asn Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile
 1105 1110 1115 1120
 Ser Ala Lys Asp Val Thr Val Asn Asn Val Thr Ser His Lys Thr
 1125 1130 1135

Ile Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr
 1140 1145 1150
 Thr Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly
 1155 1160 1165
 Thr Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr
 1170 1175 1180
 Glu Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly
 1185 1190 1195 1200
 Thr Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu
 1205 1210 1215
 Ser Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys
 1220 1225 1230
 Val Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly
 1235 1240 1245
 Ala Leu Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn
 1250 1255 1260
 Ala Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser
 1265 1270 1275 1280
 Ser Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val
 1285 1290 1295
 Gly Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys
 1300 1305 1310
 Leu Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr
 1315 1320 1325
 Thr Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr
 1330 1335 1340
 Val Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala
 1345 1350 1355 1360
 Lys Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly
 1365 1370 1375
 Lys Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln
 1380 1385 1390
 Thr Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala
 1395 1400 1405
 Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp
 1410 1415 1420
 Ser Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp
 1425 1430 1435 1440
 Ala Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala
 1445 1450 1455
 Thr Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser
 1460 1465 1470
 Val Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile
 1475 1480 1485

102

Ser Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp
 1490 1495 1500
 Val Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu
 1505 1510 1515 1520
 Ala Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg
 1525 1530 1535
 Glu Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro
 1540 1545 1550
 Asn Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro
 1555 1560 1565
 Ser Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly
 1570 1575 1580
 Asn Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595 1600

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val Asp Glu Val Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp
 1 5 10 15
 Leu Ser Asp Glu Glu Arg Glu Ala Leu Ala Lys Leu Gly
 20 25

SEQUENCE LISTING

<110> Barenkamp, Stephen J.

<120> HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
HAEMOPHILUS

<130>

<140>

<141>

<150> 09/155,614

<151> 1998-09-30

<150> 08/617,697

<151> 1996-04-01

<150> PCT/US97/04707

<151> 1997-04-01

<160> 11

<170> PatentIn Ver. 2.1

<210> 1

<211> 5116

<212> DNA

<213> Haemophilus influenzae

<400> 1

acagcgttct cttataacta gtacaaaccc acaataaaat atgacaaaca acaattacaa 60
cacctttttt gcagtctata tgcaaatatt taaaaaaaata gtataaatcc gccatataaa 120
atggtataat ctttcatctt tcacatttca tctttcatct ttcacatttc atctttcatc 180
tttcacatctt catcttcat ctttcacat tcatctttca tctttcatct ttcacatttc 240
acatgcctg atgaaccgag ggaaggaggag gaggggcaag aatgaagagg gagctgaacg 300
aacgcaaatg ataaagtaat ttaattgttc aactaacctt aggagaaaat atgaacaacg 360
tatatcgctc caaattcagc aaacgcctga atgcttttgt tgctgtgtc gaattggcac 420
ggggttgtga ccattccaca gaaaaaggca gcgaaaaacc tgctcgcatg aaagtgcgtc 480
acttagcggtt aaagccactt tccgctatgt tactatctt aggtgtaca tctattccac 540
aactctttt agcaagcgcc ttacaaggaa tggatgttagt acacggcaca gccactatgc 600
aagtagatgg taataaaacc attatccgca acagtgttga cgatatcatt aattggaaac 660
aatttaacat cgacccaaat gaaatggtgc agtttttaca agaaaacaac aactccgccc 720
tattcaaccc tgttacatct aacccaaatct cccaaattaaa agggatttta gattctaacg 780
gacaagtctt ttaatcaac cccaaatggta tcacaatagg taaagacgca attattaaca 840
ctaatggctt tacggcttct acgctagaca tttctaacga aaacatcaag ggcgttaatt 900
tcacccctcg acaaaccaaa gataaaagcgc tcgctgaat tggatcac ggtttaatta 960
ctgtcggtaa agacggcagt gtaaatctt ttggggcaaa agtggaaaac gagggtgtga 1020
ttagcgtaaa tggtggcagc atttcttac tcgcaggccca aaaaatcacc atcagcgata 1080
taataaaaccc aaccattact tacagcattt cccgcctga aatgaagcg gtcaatctgg 1140
gcgatatttt tgccaaaggc ggtacattt atgtccgtgc tgccactatt cgaaaccaag 1200
gtaaactttc tgctgattct gtaagcaaag ataaaagccg caatattgtt ctttccgcca 1260
aagagggtga agcggaaatt ggcgtgtaa tttccgctca aaatcagcaa gctaaaggcg 1320
gcaagctgat gattacaggg gataaaagtca cattaaaaac aggtgcagtt atcgacctt 1380
caggtaaaga agggggagaa acttaccttgc gcggtgacga ggcggcgaa ggtaaaaagg 1440
gcattcaatt agcaaagaaa acctcttttag aaaaaggctc aaccatcaat gtatcaggca 1500
aagaaaaagg cggacgcgtc attgtgtggg gcgatattgc gttaaattgac ggcataatatta 1560
acgctcaagg tagtgggtat atcgctaaaa ccgggtggtt tggagacg tcggggcatg 1620
atttattcat caaagacaat gcaattgttg acgccaaaga gtgggttta gacccggata 1680
atgtatctat taatgcagaa acagcaggac gcagcaatac ttcagaagac gatgaataaca 1740
cgggatccgg gaatagtgcg agcaccacca aacgaaacaa agaaaagaca acattaacaa 1800

acacaactct tgagagtata ctaaaaaaag gtaccttgc taacatcact gctaatac 1860
 gcatctatgt caatagctcc attaattttt ccaatggcag ctttaactctt tggagtgagg 1920
 gtcggagcgg tggcggcggt gagattaaca acgatattac caccgggtat gataccagag 1980
 gtgcaaactt aacaatttac tcaggcggct gggttgtatg tcataaaaaat atctcaactcg 2040
 gggcgcaggtaa acataaaac attacagctt aacaagatgat cgccttgag aaaggaagca 2100
 accaagtcat tacaggctaa gggactatta cctcaggcctt tcaaaaaaggt ttagattta 2160
 ataatgtctc tcttaaacggc actggcagcg gactgcaatt caccactaaa agaaccata 2220
 aatacgttat cacaataaaa tttgaaggaa cttaaatat ttcaggggaa gtgaacatct 2280
 caatggttt acctaaaaat gaaagtggat atgataaaatt caaaggacgc acttactgg 2340
 atttaacctc cttaaatgtt tccgagatgt gcgagtttaa cctcactatt gactccagag 2400
 gaagcgatag tgcaggcaca cttacccagc ctataattt aaacggtata tcattcaaca 2460
 aagacactac cttaatgtt gaacgaaatg caaggtcaa ctttgacatc aaggcaccaa 2520
 tagggataaa taagtattct agttgaatt acgcatcatt taatggaaac atttcagttt 2580
 cgggggggg gagtgttgc ttcacacttc tcgcctcatc ctctaacgtc caaacccccc 2640
 gttagttat aaattctaaa tacttaatg tttcaacagg gtcaagttt agatttaaa 2700
 cttcaggctc aacaaaaact ggcttctaa tagagaaaatg ttaacttta aatgccaccg 2760
 gaggcaacat aacactttt cagttgaag gcaccgttgc aatgatttttgg 2820
 tagccaaaaaa aaacataacc tttgaaggag gtaacatcac ctttggctcc agggaaagccg 2880
 taacagaaaat cgaaggcaat gttactatca ataacaacgc taacgtcaact cttatcggtt 2940
 cggattttga caaccatcaa aaacctttaa ctattaaaaa agatgtcatc attaatacg 3000
 gcaaccttac cgctggggc aatattgtca atatagccgg aaatcttacc gttgaaagta 3060
 acgctaattt caaagctatc acaaatttca cttaatgtt aggcggcttgc tttgacaaca 3120
 aaggcaatttca aatattttcc attgccaatgg gggggcttgc cttaaagac attgataatt 3180
 ccaagaattt aagcatcacc accaacttca gtcactt cgcacttata attagccgg 3240
 atataaccaaa taaaacggt gatttaataa ttacgaacga aggttagtgc actgaaatgc 3300
 aaattggcg gcatgtctcg caaaaagaag gtaatctcac gattttcttgc gacaaaatca 3360
 atattaccaaa acagataaca atcaaggcag gtgttgcatttgg ggagaatttcc gattcagacg 3420
 cgacaaacaa tgccaaatcta accattaaaaa ccaaagaattt gaaattaacg caagacctaa 3480
 atatttcagg tttcaataaa gcagagatta cagctaaaga tggtagtgcatttgc 3540
 gtaacaccaaa tagtgcgtat ggtactaatg ccaaaaaaggt aacctttaac caggttaaag 3600
 attcaaaaat ctctgcgtac ggtcacaagg tgacactaca cagcaagtg gaaacatccg 3660
 gtagtaataa caacactgaa gatagcagtgc acaataatgc cggcttaact atcgatgc 3720
 aaaatgtac agttaacaaac aatattactt ctcacaaacg agtgcgtatc tctgcgcacaa 3780
 gtggagaaat taccactaaa acaggtacaa ccattaacgc aaccactggat aacgtggaga 3840
 taaccgctca aacaggtatg atccttaggtt gaaattggatc cagctctggc tctgtacac 3900
 ttactgcac accatttttgc gcaatatttgc gggcaacacc gttactgtt 3960
 ctgcaatag cgggtcatta accattttgg caggctctac aattaaaggaa accgagatgt 4020
 taaccacttc aagtcaatca ggcgatatcg gcggtacgtat ttctggggc acagtagagg 4080
 tttaagcaac cggaaatgtt accactcaat ccaattcaaa aattaaaggaa acaacaggcg 4140
 aggctaacgt aacaagtgc acaggtacaa ttgggtgtac gattttccggtaaatacggtaa 4200
 attttacggc aaacgcgttgc gatttaacag ttggggatgg cgcagaaattt aatgcgcacag 4260
 aaggagctgc aaccttaact acatcatcg gcaaaatttac taccgaagct agttcacaca 4320
 ttacttcagc caagggtcag gtaaatctt cagctcaggatc tggtagtgc ttgcaggaa 4380
 ttaatgccgc caatgtgaca ctaaaatacta caggcactt aactaccgtg aagggttcaa 4440
 acattaaatgc aaccagcggtt accttggtaa ttaacgcacaa agacgtgc gtaaatggcg 4500
 cagcattttggg taaccacaca gtggtaatg caaccaacgc aatggctcc ggcagcgtaa 4560
 tcgcgacaac ctcaagcaga gtgaacatca ctggggattt aatcacaata aatggattaa 4620
 atatcatttc aaaaaacggt ataaacaccg tactgtttaa aggcgttata attgatgtga 4680
 aatacattca accgggtata gcaagcgtatc atgaagttat tgaagcgaaa cgcatttc 4740
 agaaggtaaa agatttatct gatgaagaaa gagaagcggtt agctaaactt ggagtaatgt 4800
 ctgtacgtttt tattgagcca aataatacaat ttacagtgc tacacaaaat gaatttgca 4860
 ccagaccatt aagtgcgatata gtcgttgc aaggcaggcgtt gtcgttgcata aacagtgc 4920
 ggcgcgacgggt gtcgttgcata atcgcgtata acggcgggtt gtcgttgcata attgacaagg 4980
 tagatttcat cctgcataatgc agtcaattttgc ttttgcattt atttactgtg tgggtttaaag 5040
 ttcaatgcgg gctttacca tcttgcataatgc aattacggat aatacaataa agtattttta 5100
 acaggttattt attatgc 5116

<210> 2
 <211> 1536
 <212> PRT
 <213> Haemophilus influenzae

<400> 2
 Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95

Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr
 450 455 460
 Gly Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe
 485 490 495
 Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly
 515 520 525
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly
 530 535 540
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn
 545 550 555 560
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp
 565 570 575
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr
 580 585 590

Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys
 610 615 620
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys
 625 630 635 640
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys
 645 650 655
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu
 660 665 670
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala
 675 680 685
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys
 690 695 700
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile
 705 710 715 720
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser
 725 730 735
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Ser Val Asp Phe Thr
 740 745 750
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn
 755 760 765
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr
 770 775 780
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu
 785 790 795 800
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp
 805 810 815
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu
 820 825 830
 Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu
 835 840 845
 Gly Asn Val Thr Ile Asn Asn Ala Asn Val Thr Leu Ile Gly Ser
 850 855 860
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile
 865 870 875 880
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala
 885 890 895
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn
 900 905 910

Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn
 915 920 925
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser
 930 935 940
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Ser Thr Tyr Arg Thr Ile
 945 950 955 960
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn
 965 970 975
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys
 980 985 990
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln
 995 1000 1005
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala
 1010 1015 1020
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr
 1025 1030 1035 1040
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys
 1045 1050 1055
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr
 1060 1065 1070
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly
 1090 1095 1100
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr
 1105 1110 1115 1120
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Asn Ile Thr Ser His Lys
 1125 1130 1135
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly
 1140 1145 1150
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr
 1155 1160 1165
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu
 1170 1175 1180
 Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr
 1185 1190 1195 1200
 Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser
 1205 1210 1215

Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp
 1220 1225 1230
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu
 1235 1240 1245
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu
 1250 1255 1260
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly
 1265 1270 1275 1280
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn
 1285 1290 1295
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser
 1300 1305 1310
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys
 1315 1320 1325
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile
 1330 1335 1340
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val
 1345 1350 1355 1360
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala
 1365 1370 1375
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val
 1380 1385 1390
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser
 1395 1400 1405
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn
 1410 1415 1420
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys
 1425 1430 1435 1440
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Glu Val
 1445 1450 1455
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu
 1460 1465 1470
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile
 1475 1480 1485
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr
 1490 1495 1500
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser
 1505 1510 1515 1520
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg
 1525 1530 1535

<210> 3
<211> 4937
<212> DNA
<213> Haemophilus influenzae

<400> 3
taaatataca agataataaa aataaatcaa gatTTTgtg atgacaaaca acaattacaa 60
cacTTTTT gcagtctata tgcaaatatt taaaaaaat agtataaatac cgccatataa 120
aatggtataa tctttcatct ttcatcttta atcttcatc tttcatctt catcttcat 180
ctttcatctt tcacatTTca tctttcatct ttcatcttca atcttcatc tttcatctt 240
cacatgaaat gatgaacca gggaaaggag ggagggcaa gaatgaagag ggagctgaa 300
gaacgcaat gataaagtaa tttaattgtt caactaactt taggaaaaaa tatgaacaag 360
atatacgta tcaaatttcg caaacgcctg aatgcttgg ttgctgtgc tgaattggca 420
cggggtgtg accattccac agaaaaaggc ttccgcctatg ttactatctt taggtgtaa 480
cacttagcgt taaagccact ttccgcctatg ttactatctt taggtgtaa 540
caatctgtt tagcaagccg cttacaagga atggatgtt tacacggcac agccactatg 600
caagtagatg gtaataaaac cattatccgc aacagtgtt acgctatcat taattggaaa 660
caatttaaca tcgaccaaaa tgaaatggt cagTTTTac aagaaaaacaa caactccgccc 720
gtattcaacc gtgttacatc taaccaatc tcccaattaa aagggatttt agattctaac 780
ggacaagtct tttaatcaa cccaaatggt atcacaatag gtaaagacgc aattattaac 840
actaatggct ttacggctt tacgttagac atttctaacg aaaacatcaa ggcgcgtaat 900
ttcacctcg agcaaaccaa agataaagcg ctcgctgaaa ttgtgaatca cggtttaatt 960
actgtcggt aagacggcag tgtaatctt attggggca aagtggaaaaa cgagggtgtg 1020
attagcgtaa atggggcag cattttta ctcgcaggc aaaaatcac catcagcgat 1080
ataataaacc caaccattac ttacagcatt gccgcgcctg aaaatgaagc ggtcaatctg 1140
ggcgatattt ttgccaagg cgtaacatt aatgtccgtg ctgcactat tcgaaaccaa 1200
ggtaaacttt ctgctgatc tgtaagcaaa gataaaagcg gcaatattgt tctttccgccc 1260
aaagaggggtg aagcgaaat tggcggtgtt atttccgctc aaaatcagca agctaaaggc 1320
ggcaagctga tgattacagg cgataaagtc acattaaaaa caggtgcagt tatcgacctt 1380
tcaggtaaag aagggggaga aacttacctt ggcggtgacg agcgcggcga aggtaaaaac 1440
ggcattcaat tagcaaaagaa aaccttttta gaaaaaggct caaccatcaa tggatcaggc 1500
aaagaaaaag ggggacgcgc tattttgtgg ggcgatattt cgtaatttga cggcaatatt 1560
aacgctcaag gtatgtgtt tattttttttt accgggtgtt ttgtggagac atcggggcat 1620
tattttatcca ttgacagcaa tgcaattgtt aaaacaaaaa agtgggtgtt agaccctgtat 1680
gatgtaccaa ttgaagccga agacccctt cgcaataataa cccgtataaa tgatgtatcc 1740
ccaaacaggca cgggtgaagc aagcgaccct aaaaaaaaaa gcaactcaa aacaacgcta 1800
accaatacaa ctatttcaaa ttatctgaaa aacgcctgga caatgaatata aacggcatca 1860
agaaaaactta cggtaatag ctcaatcaac atcggaaagca actccactt aattctccat 1920
agtaaaggc agcgtggcg agggttcag attgtggag atattactt taaaggccga 1980
aatttaacca ttatttctgg cggatgggtt gatgttcata aaaatattac gcttgcgtat 2040
ggtttttaa atattaccgc cgctccgtt gctttgtt gttggaaataa caaaggcacgc 2100
gacggggcaa atgctaaaat tgcgtcccgag ggcactgtt ccattacagg agagggaaaa 2160
gatttcaggc ctaacaacgt atctttaaac ggaacgggtt aaggctgtt tatttttttca 2220
tcagtgaata atttaaccca caatcttagt ggcacaattt acatatctgg gaatataaca 2280
attaacccaa ctacgagaaa gaacacctcg tattggcaaa ccagccatga ttgcactgg 2340
aacgtcagt ctcttaatct agagacaggc gcaatttttta ctttttattaa atacatttca 2400
agcaatagca aaggcttaac aacacagt agaagctctg cagggtgtt ttttaacggc 2460
gttaaatggca acatgtcatt caatctcaaa gaaggagcga aagttaattt caaattaaaa 2520
ccaaacgaga acatgaacac aagcaaaccct ttaccaattt gtttttttgc caatatcaca 2580
gccactgggtt ggggctctgt ttttttgc atatgttca accatctgg cagaggggtt 2640
gagtttttttttggat taatatctt aacggcgctt attttacattt aaattcccat 2700
gttcgcggcg atgacgcttt taaaatcaac aaagacttta ccataatgc aaccaattca 2760
aatttcagcc tcagacagac gaaagatgtt tttttatgc ggtacgcacg caatgcattc 2820
aattcaaccc acaacatatac cattctggc ggtatgtca cccttgggtgg acaaaaactca 2880

<210> 4

<211> 1477

<212> PRT

<213> *Haemophilus influenzae*

<400> 4

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Ash Ala Leu
1 5 10 15

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
20 25 30

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
35 40 45

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125
 Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400

Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro
 450 455 460
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys
 465 470 475 480
 Thr Thr Leu Thr Asn Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp
 485 490 495
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile
 500 505 510
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg
 515 520 525
 Gly Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Asn
 530 535 540
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr
 545 550 555 560
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu
 565 570 575
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala
 580 585 590
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn
 595 600 605
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser
 610 615 620
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly
 625 630 635 640
 Asn Ile Thr Ile Asn Gln Thr Arg Lys Asn Thr Ser Tyr Trp Gln
 645 650 655
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr
 660 665 670
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly
 675 680 685
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val
 690 695 700

Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe
 705 710 715 720
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile
 725 730 735
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Ser Val Phe Phe
 740 745 750
 Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser
 755 760 765
 Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val
 770 775 780
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala
 785 790 795 800
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp
 805 810 815
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu
 820 825 830
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ser Ile Thr
 835 840 845
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn
 850 855 860
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly
 865 870 875 880
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp
 885 890 895
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys
 900 905 910
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr
 915 920 925
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr
 930 935 940
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg
 945 950 955 960
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile
 965 970 975
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr
 995 1000 1005
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser
 1010 1015 1020

Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn
 1060 1065 1070
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser
 1075 1080 1085
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys
 1090 1095 1100
 Thr Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr
 1105 1110 1115 1120
 Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr
 1125 1130 1135
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr
 1140 1145 1150
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr
 1155 1160 1165
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val
 1170 1175 1180
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala
 1185 1190 1195 1200
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly
 1205 1210 1215
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu
 1220 1225 1230
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr
 1235 1240 1245
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile
 1250 1255 1260
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile
 1265 1270 1275 1280
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr
 1285 1290 1295
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu
 1300 1305 1310
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp
 1315 1320 1325

Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr
 1330 1335 1340
 Ala Ala Thr Ser Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val
 1345 1350 1355 1360
 Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu
 1365 1370 1375
 Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser
 1380 1385 1390
 Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp
 1395 1400 1405
 Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala
 1410 1415 1420
 Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn
 1425 1430 1435 1440
 Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys
 1445 1450 1455
 Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala
 1460 1465 1470
 Asp Asp Gly Gln Pro
 1475

<210> 5
 <211> 9171
 <212> DNA
 <213> Haemophilus influenzae

<400> 5
 acagcgttct cttataacta gtacaaaccc acaataaaat atgacaaaca acaattacaa 60
 cacctttttt gcagtctata tgcaaataatt ttaaaaaata gtataatcc gccatataaa 120
 atggtataat ctttcatctt tcattttca tctttcatct ttcatttttc atctttcatc 180
 tttcatcttt catctttcat ctttcatctt tcattttca tctttcatct ttcatttttc 240
 acatgaaatg atgaaccgag ggaagggagg gaggggcaag aatgaagagg gagctgaacg 300
 aacgcaaatg ataaagtaat ttaattgttc aactaacctt aggagaaaat atgaacaaga 360
 tatatcgctt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 420
 ggggttgtga ccattccaca gaaaaaggca gcgaaaaacc tgctcgatg aaagtgcgtc 480
 acttagcggtt aaagccactt tccgctatgt tactatctt aggtgtaaaca tctattccac 540
 aatctgtttt agcaagcggc ttacaaggaa tggatgttagt acacggcaca gccactatgc 600
 aagttagatgg taataaaaacc attatccgca acagtgttga cgctatcatt aattggaaac 660
 aatttaacat cgaccaaaaat gaaatggtgc agtttttaca agaaaaacaac aactccgccc 720
 tattcaaccg tggatcatct aaccaaatct cccaaataaa agggatttta gattctaacg 780
 gacaagtctt tttatcaac ccaaattggta tcacaatagg taaagacgca attattaaca 840
 ctaatggctt tacggcttct acgctagaca tttcttaacga aaacatcaag ggcgcgtatt 900
 tcacccctgca gcaaaacccaa gataaagcgc tcgctgaaat tgtgaatcac ggtttattha 960
 ctgtcggtaa agacggcagt gtaatctt ttgggtggca aatggggatggta 1020
 tttagcgtaaa tggatggcagc atttcttac tcgcaggggca aaaaatcacc atcagcgata 1080
 taataaaaccc aaccattact tacagcattt cccgcgcctga aaatgaagcg gtcaatctgg 1140
 gcgatatttt tgccaaaggc ggtaacatta atgtccgtgc tgccactatt cgaaaccaag 1200
 cttccgcca aagagggtga agcggaaatt ggcgggtgaa tttccgctca aaatcagcaa 1260
 gctaaaggcg gcaagctgat gattacagggc gataaaaac aggtgcagg 1320

atcgaccctt caggtaaaga agggggagaa acttaccttgcgggtacga ggcggcgc 1380
ggtaaaaacg gcattcaatt agaaaagaaa acctctttag aaaaaggctc aaccatcaat 1440
gtatcaggca aagaaaaagg cgacgcgcattgtgtgg gcatattgc ttattgc 1500
ggcaatatta acgtcaagg tagtggtagt atcgtaaaa cgggtggtt tggagacg 1560
tcggggcatg atttattcat caaagacaat gcaattgttgcgcaaaaga gtgggttta 1620
gaccggata atgtatctat taatgcagaa acagcaggac gcaatatac ttcaagac 1680
gatgaataca cgggatccgg gaatagtgc agcaccctaa aacgaaacaa agaaaagaca 1740
acattaacaa acacaactct tgagagtata ctaaaaaag gtacccctgt taacatca 1800
gctaatac acatctatgt caatagctcc attaatttat ccaatggcag cttaactctt 1860
tggagtgagg gtcggagcgg tggccgcgtt gagattaaca acgtattac caccgggtat 1920
gataccagag gtcaaaactt aacaatttac tcaggcgcgtt ggggtatgt tcataaaaat 1980
atctcactcg gggcgcaggaa accaagtcat tacaggtca gggacttata cctcaggc 2040
aaaggaagca accaacttataatgtctc tctaaacggc actggcagcg gactgcaatt caccactaa 2100
tttagattta ataatgtctc cacaataaaa tttgaaggca ctttaataat ttcagggaaa 2160
agaaccaata aatacgctat acctaaaaat gaaagtggat atgataaatt caaaggacgc 2220
gtgaacatct caatggttt gaaagtggat atgataaatt caaaggacgc cctca 2280
acttacttga atttaacctc gaaagtggat atgataaatt caaaggacgc cctca 2340
gactccagag gaagcgatag tgcaggcaca cttacccagc cttataatt aaacgggtata 2400
tcattcaaca aagacactac ctttaatgtt gaacgaaatg caaggtcaa ctttgacatc 2460
aaggcaccaa tagggataaa taagtattct agttgaatt acgtcatcatt taatggaaac 2520
attcagttt cggggggggg gtagtggat ttcacacttc tgccttc tcttaacgtc 2580
caaaccctcg gttagttat aaattctaaa tacttaatg tttcaacagg gtcaagtta 2640
agattaaaaa cttcaggcgc aacaaaaact ggctctcaa tagagaaaat ttaacttta 2700
aatgccacccg gaggcaacat aacacttttgaagtggat atgattggc aatgattgg 2760
aaaggcattt tagccaaaaaa aaacataacc tttgaaggag gtaagatgag gtttggctcc 2820
aggaaagccg taacagaaat cgaaggcaat gttactatca ataacaacgc taacgtcact 2880
cttacggtt cggattttga caaccatcaa aaacctttaa ctattaaaaa agatgtcatc 2940
attaatagcg gcaaccttac cgctggggc aatattgtca atatagccgg aatcttacc 3000
gttgaagta acgtaattt caagctatc acaaatttca ctttaatgtt aggcggctt 3060
tttgacaaaca aaggcaatttca aatatttcc attgccaag gggggctcg ctttaaagac 3120
attgataattt ccaagaatttca aagcatcacc accaacttca gctccactt ccgcactt 3180
ataagcggca atataaccaa taaaaacggt gatttaataat ttagaacaagc aggtatgt 3240
actgaaatgc aaatttggcg ccatggatccaa aacatggcgtt gtttttttcc 3300
gacaaaatca atattaccaa agacataaca atcaaggcgtt gtttttttcc 3360
gattcagacg cgacaaacaa tgccaaatcta accattaaaaa ccaaagaattt gaaattaacg 3420
caagacctaa atatttcagg ttaacttatttgcgtt gtttttttcc 3480
ttaacttatttgcgtt gtaacaccaa cggatccaaatca aacatggcgtt gtttttttcc 3540
caggtaaag attcaaaaat tctctgttgcgtt ggttgcataagg tgacactaca cagcaaagt 3600
gaaacatccg gttagtataaa caacacttgcgtt gtttttttcc 3660
atcgatgcaaa aaaaatgtac taccactaa acaggtacaa cttccatggcgtt gtttttttcc 3720
tctgcacaa gttggagaaat aacaggtatgtt gtttttttcc 3780
aacgtggaga taaccgctca ttacttgcacatcg gtttttttcc 3840
tctgttaacac ttacttgcacatcg gtttttttcc 3900
gttactgttta ctgcaatag cggatccaaatca aacatggcgtt gtttttttcc 3960
accgagatgt taaccacttgcgtt gtttttttcc 4020
acagtagagg ttaaagcaac cggatccaaatca aacatggcgtt gtttttttcc 4080
acaacaggcg aggttaacgt aacaagtgcgtt gtttttttcc 4140
aatacggtaa atgttacggc aaacgcttgcgtt gtttttttcc 4200
aatgcgacag aaggagctgc aaccttaacttgcgtt gtttttttcc 4260
agttcacaca ttacttgcgtt caagggtcgtt gtttttttcc 4320
gcaggaaagta ttaatgcgcgtt gtttttttcc 4380
aagggttcaaa acattaaatgc aaccagcggtt gtttttttcc 4440
ctaaatggcg cagcattggg taaccacaca gtttttttcc 4500
ggcagcgttaa tcgcgacaac ctcaaggcgtt gtttttttcc 4560
aatggattaa atatcatttgcgtt gtttttttcc 4620
attgtatgttca aatacatttgcgtt gtttttttcc 4680
cgcaccccttgcgtt agaaggtaaaatca agtgcgtt gtttttttcc 4740
ggcgttaagtttgcgtt tattgagccaaataatca aacatggcgtt gtttttttcc 4800
qaatttgcgtt ccaagaccatttgcgtt aagtcgaaataatca aacatggcgtt gtttttttcc 4860

aacagtgatg ggcgcacggt gtgcgttaat atcgctgata acgggcggt acggtcagta 4920
attgacaagg tagatttcat cctgcaatga agtcattttt ttttctgtatt atttactgtg 4980
tgggttaaag ttctgatcg gctttaccca tcttgtaaaa aattacggag aataacaataa 5040
agtatttttta acaggttattt attatgaaaaa atataaaaag cagataaaaa ctcagtgcaa 5100
tatcagtattt gcttggccctg gcttcttcat cattgtatgc agaagaagcg ttttttagtaa 5160
aaggcttca gttatctgtt gcacttggaa cttaagtga agacgccccaa ctgtctgttag 5220
caaaaatctt atctaaatac caaggctcgc aaactttaac aaacctaataa acagcacagc 5280
ttgaattaca ggctgtgcta gataagattt agccaaataa gtttgcgtg atattgccac 5340
aacaaaaccat tacggatggc aatattatgt ttgagctagt ctcgaaatca gccgcagaaaa 5400
gccaagtttt ttataaggcg agccagggtt atagtgaaga aaatatcgct cgtagcctgc 5460
catcttggaa acaaggaaaaa gtgtatgaag atggctgtca gtggctcgat ttgcgtgaat 5520
tcaatatggc aaaagaaaaat ccacttaaaag tcactcgctg gcattacgag ttaaacccta 5580
aaaacaaaac ctctgatttg gtatgtcag gttttcgcc ttttgc当地 acgcgttagct 5640
ttgtttccta tgataatttc ggcccaaggg agttaacta tcaacgtgt agtctaggtt 5700
ttgttaatgc caatttgacc ggacatgatg atgttataaa tctaaacgca ttgaccaatg 5760
taaaagcacc atcaaataatc tatgcgttag gcataggata tacttatccg ttttatgata 5820
aacaccaatc ctaaagtctt tataccagca tgagttatgc tgattttaat gatatcgacg 5880
gcttaccaag tgcgatttaat cgtaaattat caaaagggttca atctatctt gcaatctga 5940
aatggagttt ttatctcccg acatttaacc ttggatggaa agaccagttt aaaatttaatt 6000
taggctacaa ctaccgcattt attaatcaaa catccgagttt aaacaccctg ggtcaacgaa 6060
agaaaaaattt tgcaatgtca ggcgtaaatg caggcatttga tggacatatac caatttaccc 6120
ctaaaacaat cttaatattt gatttaactc atcattattt cgcgagtaaa ttaccaggct 6180
cttttggat ggagcgcattt ggccaaacat ttaatcgacg ctatcacattt agcacagcca 6240
gttttagggtt gagtcaagag tttgtcaag gttggcattt tagcagtcgaa ttatcggttc 6300
agtttactctt acaagatata agtagcatag atttatttctc tgtaacaggtt acttatggcg 6360
tcagaggctt taaatacggc ggtcaatgtt gtgagcgcgg tcttgcgtt cgtatgtat 6420
taagtatgcc aaaatacacc cgcttcaaa tcagccctta tgcttttat gatgcaggtc 6480
agttccgttta taatagcgaa aatgtaaaaa cttacggcga agatatgcac acggatccct 6540
ctgcgggtttt aggcattttt accttcctt cacaactt aagcttagat gctttgttg 6600
ctcgctcgctt tgcaatgtcc aatagtgaca atttgcatttgg caacaaaaaa cgcacaagct 6660
cacctacaac cttctgggtt agattaacat tcagtttctt accctgaaat ttaatcaact 6720
ggtaagcggtt ccgcctacca gtttataactt atatgtttt cccgcattt tacagtctat 6780
acgcacccctt gtttcatcc ttatatatca aacaaactt gcaaaaccaag caaaccaagc 6840
aaaccaagca aaccaagcaaa accaagcaaa ccaagcaac caagcaaacc aagcaaacca 6900
agcaaaccaaa gcaaaaccaag caaaccaagc aaaccaagca atgctaaaaa acaatttata 6960
tgataaaacta aacataactc cataccatgg caatacaagg gatttaataa tatgacaaaa 7020
gaaaattttac aaagtgttcc acaaaatacg accgttccac ttgttagaattc aaacaacgac 7080
caaacttccc tgcaatataact taaacaacca cccaaaccca accttattacg cctggaaccaa 7140
catgtcgccca aaaaagatta tgagcttgc tggcgcaat taatggcgat ttggaaaaaa 7200
atggacgctt atttggagg cggttccatgatttgcatttgc tcagctggca 7260
tatcttcccg aaaaactact aatttattt gcaactcgatc tgcgtatgc aattacaaca 7320
ctctttccg accccgaattt ggcaatttcc gaagaagggg cattaaagat gattagcctg 7380
caacgctggat tgacgctgtt ttttgcctt tccccctacg ttaacgcaga ccatattctc 7440
aataaaatata atatcaaccc agatccgaa ggtggcttc atttaccaac agacaactct 7500
tctattgcta aatttctgtat ttttactt cccgaatcca atgtcaatat gagtttagat 7560
gcgttatggg cagggatca acaactttgtt gtttgcattt gtttgcgtt gcagtcttca 7620
cgtttatttgc tgcgttcat aaaagagcggtt tggtttaca gtggttccct 7680
aaaaaactcg ccgaaatttc taatttagat gaattgcctt caaatatcc tcatgtatgt 7740
tatatgcact gcagttatgtt ttttagaaaaa aacaagcactt atgttaagcg tccattaaac 7800
gaacttgcac gcaagcatat cctcacgcaaa ggatggcaag accgcacccctt taccacctt 7860
ggtaaaaagg acggcaaccc ttttgcgttgc tgcgttcat aaaagagcggtt tggtttaca gtggttccct 7920
tcgatttacgc gcacgcattt aacttcaatg atttgcgttgc gagaaaaaattt ctattttagtc 7980
ggcttaggcc atgagggcgat tgataacataa ggatggcaag ttttgcgtt gtttgcgtt 8040
atcagtagca ataataatggagactt gtttgcgttgc ttttgcgtt gtaaaacagtg cgaaactttc 8100
caacccgcacg ttttgcgttgc ttttgcgttgc ttttgcgtt gtaaaacagtg cgaaactttc 8160
aacactcgcc ttggccctt tcaagctgtt gcttggccat atccttgcac tacgcatttc 8220
gaatttatttgc attatgttcat cgtagaagat gattatgtgg gcaatgttgc ttgttttagt 8280
gaaaccctt tacgcttacc caaagatgccc ctacccatgc taccatctgc actcgccccaa 8340
caaaaagtgg attatgtact caggaaaaac cctgaagtagt tcaatatcggtt tattggcgat 8400

accacaatga aattaaaccc tgaattttt ctaacattgc aagaaaatcg agataaaagct 8460
 aaagtcaaaa tacatttca ttcgcactt ggacaatcaa caggcttgc acacccttat 8520
 gtcaaatggt ttatcgaaag ctatttaggt gacgatgcca ctgcacatcc ccacgcacct 8580
 tatacagatt atctggcaat attgcgtgat tgcgatatgc tactaaatcc gtttccttcc 8640
 ggtaatacta acggcataat tgatatggtt acatttaggt tagttgggt atgcaaaaacg 8700
 ggggatgaag tacatgaaca tattgatgaa ggtctgttta aacgctttagg actaccagaa 8760
 tggctgatag ccgacacacg agaaaacatat attgaatgtg ctttgcgtct agcagaaaaac 8820
 catcaagaac gccttgaact ccgtcggtac atcatagaaa acaacggctt acaaaaagctt 8880
 tttacaggcg accctcgcc attgggcaaa atactgctta agaaaacaaa tgaatggaaag 8940
 cggaaagcact tgagtaaaaa ataacggttt tttaaagtaa aagtgcgggtt aattttcaaa 9000
 gcgtttaaa aacctctcaa aaatcaaccg cacttttac tttataacgc tcccgcgcc 9060
 tgacagttt tctctttttt aaaaataccctaaaattgtg gcaataatgtt ggtatcaaa 9120
 ttcaattgtt gatacggcaa actaaagacg gcgcgttctt cgccagtcat c 9171

<210> 6

<211> 9323

<212> DNA

<213> Haemophilus influenzae

<400> 6

cgccacttca attttggatt gttgaaattc aactaaccaa aaagtgcggt taaaatctgt 60
 ggagaaaata ggtttagtg aagaacgagg taattgttca aaaggataaa gctctcttaa 120
 ttggcattt gttggcggtt cttttcggt taatagtaaa ttatattctg gacgactatg 180
 caatccacca acaactttac cgttgggtt aacgcttaat gtaagttctt gctcttcttg 240
 gcaataacgt aatcccattt tttgttttagc aagaaaaatga tcgggataat cataataggt 300
 gttgcccaaa aataaaatttt gatgttctaa aatcataaat tttcaagat attgtggcaa 360
 ttcaataacctt atttgtggcg aaatcgccaa ttttaatttca atttcttgcataatatt 420
 tcccaactcaa atcaacttgcg taaatataca agataataaa aataaatcaa gatttttgcgt 480
 atgacaacaca acaattacaa cacctttttt gcagttctata tgcaaatatt ttaaaaaaaat 540
 agtataaattc cgccatataaa aatgttataa tctttcatct ttcatctttc atctttcatc 600
 tttcatcttt catctttcatctt tcatctttca tctttcatct ttcatctttc 660
 atctttcatc tttcatcttt cacatgaaat gatgttgcg gggaaaggag ggaggggcaa 720
 gaatgaagag ggagctgaac gaacgcaaat gataaaagtaa tttaaatgtt caactaacct 780
 taggagaaaa tatgaacaag atatatcgtc tcaaatttcg caaacgcctg aatgttttg 840
 ttgtgtgtc tgaattggca cgggttgcgtt accattccac agaaaaaggc agcggaaaaac 900
 ctgctcgcat gaaagtgcgt cacttagcgt taaagccact ttccgctatg ttactatctt 960
 taggtgtaac atctatttca caatctgttt tagcaagccg caatttaaca tcgaccaaaa 1020
 tgaatgggtc cagttttac aagaaaacaa gtaataaaac cattatccgc aacagtgtt 1080
 acgtatcat taattggaaa caatttaaca tcgaccaaaa tgaatgggtc cagttttac 1140
 aagaaaacaa caactccgccc gtattcaacc gtgttacatc taaccaaattc tcccaattaa 1200
 aagggatttt agattctaac ggacaagtct ttttaatcaa cccaaatggt atcacaatag 1260
 gtaaagacgc aattttaaac actaatggct ttacggctc tacgttagac atttctaaacg 1320
 aaaacatcaa ggccgtat ttcacccctcg agcaaaccaa agataaaggc ctgcgtgaaa 1380
 ttgtgaatca cgggtttaattt actgtcggtt aagacggcag tgtaaatctt attgggtggca 1440
 aagtgaaaaaa cgagggtgtt attagcgtaa atgggtggcag catttcttta ctgcgaggcc 1500
 aaaaaatcac catcagcgat ataataaacc caaccattac ttacagcatt gccgcgcctg 1560
 aaaaatgaagc ggtcaatctg ggcgtatattt ttgccaaagg cggtaacatt aatgtccgtg 1620
 ctgccactat tcgaaaccaa ggttaacttt ctgctgattc tgtaagcaaa gataaaaggcg 1680
 gcaatattgt tctttccccc aaagagggtt aagcggaaat tggcgtgtt atttccgctc 1740
 aaaatcagca agctaaaggc ggcaagctga tgataaaagtc cgataaaagtc acattaaaaa 1800
 caggtgcagt tatacgacatt tcaggtaaag aagggggaga aacttacattt ggcgggtgacg 1860
 agcgcggcga aggtaaaaac ggcattcaat tagcaaagaa aacctttaa gaaaaaggct 1920
 caaccatcaa tttatcggtt aaagaaaaag gccgcgcgc tattgtgtgg ggcgtatattg 1980
 cgttaatttca cggcaatattt aacgctcaag gtatgtgtt gatcgctaaa accgggtgtt 2040
 ttgtggagac atcgggccat tatttattca ttgacagcaa tgcaattgtt aaaaacaaaag 2100
 agtgggttgcgat agaccctgtat gatgtaaacaa ttgaaaggccgaa agacccctt cgcaataata 2160
 ccggatataaa tttatcggtt aacacaggca ccgggtgtt gatcgctaaa accgggtgtt 2220
 gcaactcaa aacaacgcata accaatacaa ctatttcaaa ttatctgaaa aacgccttggaa 2280

caatgaataat aacggcatca agaaaactta ccgttaatag ctaaatcaac atcggagaagc 2340
actcccaactt aattctccat agtaaaggcgc agcgtggcgg aggcgttcag attgatggag 2400
atattacttc taaaggcggaa aatttaacca tttattctgg cggatgggtt gatgttcata 2460
aaaatattac gcttgatcag gttttttaa atattaccgc cgcttccgta gctttgaag 2520
gtggaaataaa caaagcacgc gacgcggcaa atgctaaaat tgtcgcccag ggcactgtaa 2580
ccattacagg agagggaaaa gatttcaggg ctaacaacgt atcttaaac ggaacgggta 2640
aaggctctgaa ttcatttca tcagtgataa atttaacccaa caatcttagt ggcacaatta 2700
acatatctgg gaatataaca attaaccaaa ctacgagaaaa gaacacctcg tattggcaaa 2760
ccagccatga ttgcactgg aacgtcagt ctctaatct agagacaggc gcaaatttta 2820
cctttattaa atacatttca agcaatagca aaggcttaac aacacagtat agaagctctg 2880
caggggtgaa ttttaacgc gtaaatggca acatgtcatt caatctcaa gaaggagcga 2940
aagtttaattt caaattaaaaa ccaaacgaga acatgaacac aagcaaaactt ttaccaattc 3000
ggtttttagc caatatcaca gccactgggt ggggtctgt tttttttagt atatatgcca 3060
accattctgg cagaggggct gagttaaaaa tgagtgaaat taatatctt aacggcgcta 3120
attttacccaaattccat gttcgcggcg atgacgctt taaaatcaac aaagacttaa 3180
ccataaaatgc aaccaattca aatttcagcc tcagacagac gaaagatgtat ttttatgacg 3240
ggtacgcacg caatgccatc aattcaacct acaacatatc cattctggc gtaatgtca 3300
cccttggggc acaaaaactca agcagcagca ttacggggaa tattactatc gagaaggcag 3360
caaatgttac gctagaagcc aataacgccc ctaatcagca aaacataagg gatagagtt 3420
taaaacttgg cagcttgctc gttaatggga gtttaagttt aactggcga aatgcagata 3480
ttaaaggcaa tctactatt tcagaaagcg ccactttaa aggaaagact agagataacc 3540
taaatatcac cgcaattttt accaataatg gcactgccga aattaatata acacaaggag 3600
tggtaaaact tggcaatgtt accaatgatg tgatgttaaa cattaccact cacgtaaac 3660
gcaaccaaag aagcatcatc ggcggagata taatcaacaa aaaaggaagc tttaatatta 3720
cagacagtaa taatgtatgt gaaatccaaa ttggcggca tatctcgcaaa aaagaaggca 3780
acctcacat ttctccgat aaaattaata tcaccaaaaca gataacaatc aaaaagggtt 3840
ttgatggaga ggactctagt tcagatgcga caagtaatgc caacctaact attaaaacca 3900
aagaattgaa attgacagaa gacctaagta tttcaggttt caataaagca gagattacag 3960
ccaaagatgg tagagattt actattggca acagtaatga cggttaacagc ggtgccgaag 4020
ccaaaacagt aacttttaac aatgttaaag attcaaaaat ctctgctgac ggtcacaatg 4080
tgacactaaa tagcaaagtg aaaacatcta gcagcaatgg cgacgtgaa agcaatagcg 4140
acaacgatac cggcttaact attactgcaaaaatgtaga agtaaacaata gatattactt 4200
ctctaaaaac agtaaatatc accgcgtcgg aaaaggttt caccacagca ggctcgacca 4260
ttaacgcaac aaatggcaaa gcaagtatta caaccaaaac aggtgatatc agcgtacga 4320
tttccggtaa cacggtaagt gttagcgcga ctggtgattt aaccactaaa tccgctcaa 4380
aaattgaagc gaaatcggtt gaggctaattg taacaagtgc aacaggtaca attggcggt 4440
caatttccgg taatacggta aatgttacgg caaacgctgg cgattnaaca gttggaaatg 4500
gcccggaaaat taatgcgaca gaaggagctg caacctaacc cgcacaggaa aataccttga 4560
ctactgaagc cgggttcttagc atcaatttca ctaagggtca ggttagaccc ttggctcaga 4620
atggtagcat cgcaggaaagc attaatgtctg ctaatgtgac attaaatact acaggcacct 4680
taaccaccgt ggcaggctcg gatattaaag caaccagcgg caccttgggtt attaacgca 4740
aagatgttac gcttaatgtt gatgcatttgc gtgatgtac agaagtgaat gcagtcaacg 4800
actggggatt tggtagtgcg actgcggca cctcaaggcag tggtaatatac actggggatt 4860
taaacacagt aaatgggtt aatatcattt cgaaaagatgg tagaaacact gtgcgtttaa 4920
gaggcaagga aattgagggtt aatataatcc agccagggtt agcaagtgtt gaagaagtt 4980
tggtaagcgaa acgcgtccctt gaaaaagttaa aagatttac tgatgaagaa agagaaacat 5040
tagcttaact tggtaagt gctgtacgtt ttgttgagcc aaataataca attacagtca 5100
atacacaataa tgaatttaca accagaccgt caagtcaagt gataatttctt gaaggtaagg 5160
cgtgtttctc aagtggtaat ggcgcacag tatgtaccaa ttgtgctgac gatggacagc 5220
cgttagtcaact aattgacaag gtagatttca tcctgcaatg aagtcttattt atttctgtat 5280
tatttactgt gtgggttaaa gttcagtgac ggcttaccc atcttgcataa aaattacgg 5340
gaatacaataa aagttttt aacagggtt tattatgaaa aatataaaaaa gcagattaaa 5400
actcagtgca atatcagtat tgcttggct ggcttctca tcattgtatc cagaagaagc 5460
gttttttagt aaaggcttcc agttatctgg tgcacttgaa actttaaatgt aagacgccc 5520
actgtctgtt gcaaaaatctt tatcttaataa ccaaggctcg caaactttaa caaacctaa 5580
aacagcacag cttaatttac aggctgtgtt agataagatt gagccaaata aatttgtatgt 5640
gatattgccc caacaaacca ttacggatgg caatatcatg tttgagctag tctcgaaatc 5700
agccgcagaa agccaagttt ttataaaggc gagccagggt tataatgttgaag aaaatatcgc 5760
tcgttagcctg ccatttttgc aacaaggaaaa agtgtatgaa gatggtcgtc agtgttgcga 5820

tttgcgtgaa tttaatatgg caaaagaaaa cccgcttaag gttacccgtg tacattacga 5880
 actaaaccct aaaaacaaaa cctctaattt gataattgcg ggcttctcg cttttggtaa 5940
 aacgcgtagc ttatattttt atgataattt cggcgcgaga gagtttaact accaacgtgt 6000
 aagctgggt tttgttaatg ccaatttaac tggtcatgtat gatgtgttaa ttataccagt 6060
 atgagttatg ctgattctaa tgatatcgac ggcttaccaa gtgcgattaa tcgtaaattha 6120
 tcaaaaggc aatctatctc tgcaatctg aatggagtt attatctccc aacatttaac 6180
 ctggcatgg aagaccaatt taaaattaat ttaggctaca actaccgcca tattaatcaa 6240
 acctccgcgt taaatcgctt gggtaaacg aagaaaaaat ttgcagtatc aggcgtaagt 6300
 gcaggcattg atggacatata ccaatttacc cctaaaacaa tctttatata tgatthaact 6360
 catcattatt acgcgatgaa attaccaggc tctttggaa tggagcgcatttggcaaaca 6420
 tttaatcgca gctatcacat tagcacagcc agtttagggt tgagtcaga gtttgctcaa 6480
 ggtggcatt ttagcagtc attatcaggt caatttactc tacaagatata tagcagtata 6540
 gatttattct ctgttaacagg tacttatggc gtcagaggct ttaataacgg cggtgcaagt 6600
 ggtgagcgcg gtcttgatg gctaatgaa ttaagtatgc caaaatacac cccgcttccaa 6660
 atcagccctt atgcgtttt tgatgcaggt cagttccgtt ataatacgaa aaatgctaa 6720
 actacggcg aagatatgc aacggtatcc tctgcgggtt taggcattaa aacctctcc 6780
 acacaaaact taagcctaga tgctttgtt gtcgtcgct ttgcaatgc caatagtgc 6840
 aatttgaatg gcaacaaaaa acgcacaaggc tcacctacaa ccttctgggg gagattaaca 6900
 ttcagttctt aaccctgaaa tttaatcaac tggtaagcgt tccgcctacc agtttataac 6960
 tataatgctt acccgccat ttacagtcta taggcaaccc tgttttacc cttatataatc 7020
 aaataaaacaa gctaagctga gctaagcaaa ccaagcaaac tcaagcaaggc caagtaatac 7080
 taaaaaaaaa atttataatga taaactaaag tataactccat gccatggcga tacaagggat 7140
 ttaataatata gacaaaagaa aatttgcacaa acgctccctca agatgcgacc gctttacttg 7200
 cggaaatthaag caacaatcaa actccccgtc gaatatttaa acaaccacgc aagcccagcc 7260
 tattacgctt ggaacaacat atcgcaaaaa aagattatga gtttgcgtt cgtgaattaa 7320
 tggtgattct ggaaaaaaatg gacgctaattt ttggaggcgt tcacgatattt gaatttgacg 7380
 cacccgctca gctggcatat ctacccgaaa aattactaat ttatttgcc actcgctctcg 7440
 ctaatgcaat tacaacactc tttccgacc ccgaattggc aatttctgaa gaagggcgt 7500
 taaagatgat tagcctgaa cgctgggtga cgctgattt tgccttcc ccctacgta 7560
 acgcagacca tattctcaat aaatataata tcaacccaga ttccgaagggt ggctttcatt 7620
 tagcaacaga caactcttctt attgctaaat tctgtatattt ttacttaccc gaatccaatg 7680
 tcaatatgag ttttagatgcg ttatggcag ggaatcaaca actttgtgt tcattgtgtt 7740
 ttgcgttgca gtcttcacgt tttattggta ccgcattctc gtttcataaa agagcgggtgg 7800
 ttttacagtg gtttctaaa aaactcgccg aaattgctaa tttagatgaa ttgcctgcaa 7860
 atatccttca tgatgtatata atgcactgca gttatgattt agcaaaaaac aagcacgatg 7920
 ttaagcgtcc attaaacgaa cttgtccgca agcatatctt cacgcaagga tggcaagacc 7980
 gctacctta caccttaggt aaaaaggacg gcaaaccgtt gatgatggta ctgcttgaac 8040
 attttatcc gggacatcg atttacgta cacattcaac ttcaatgattt gctgctcgag 8100
 aaaaattcta tttagtcggc tttaggccatg agggcgttga taaaataggt cgagaagtgt 8160
 ttgacgagtt ctttgaatc agtagcaata atataatgga gagactgttt tttatccgta 8220
 aacagtgcga aactttccaa cccgcagtgt tctatatgcc aagcattggc atggatatta 8280
 ccacgatttt tggtagcaac actcggtttt cccctattca agctgttagcc ctgggtcattc 8340
 ctggccactac gcattctgaa tttattgtattt atgtcatgtt agaagatgtat tttatccgta 8400
 gtgaagattt tttcagcgaa acccttttac gcttacccaa agatgcctta ctttatgtac 8460
 ctcttcactt cggcccccacaa aaagtttgcgtt tttttttttt gatgtttttt 8520
 atatcggtat tggcccttacc acaatgaaat taaaccctga atttttgttta acattgcaag 8580
 aaatcagaga taaagctaaa gtcaaaatatac attttcatttt cgcacttggc caatcaacag 8640
 gcttgacaca cccttatgtc aaatggttt tggcaatggc tttttttttt gatgtttttt 8700
 cacatccccca cgcacccattt cacgatttac tggcaatattt gctgtttttt gatgtttttt 8760
 taaatccgtt tcctttcggtt aataactaacg gcataatttgc tttttttttt gatgtttttt 8820
 ttggtgtatg caaaaacgggg gatgaagttt atgaacatata tttttttttt gatgtttttt 8880
 gcttaggactt accagaatgg ctgatagccg acacacgaga aacatataat tttttttttt gatgtttttt 8940
 tggcgttgcg agaaaaaccat caagaacgcg tttttttttt gatgtttttt 9000
 acggcttaca aaagctttt acaggcgcacc ctcgtttttt gggcaaaaata tttttttttt gatgtttttt 9060
 aaacaaatgaa atggaagccg aagcaacttgc gttttttttt aacggttttt 9120
 tggcgttataat tttttttttt tttttttttt tttttttttt tttttttttt 9180
 ataaacgatcc cgcacccatca cagttttatca gcctcccgcc ataaaactcc gcctttttt 9240
 gcccggatattt tagccaaaaac tggcagaaat taaaggctaa aatcaccaaaa ttgcaccacca 9300
 aaatcaccacca taccaccaaaa aaa 9323

```
<210> 7
<211> 4794
<212> DNA
<213> Haemophilus influenzae
```

tctgattcaa gtgaggcaga aaatgctaac ctaactattc aaaccaaaga gttaaaattg 3120
 gcaggagacc taaatattc aggcttaat aaagcagaaa ttacagctaa aaatggcagt 3180
 gatttaacta ttggcaatgc tagcgggtgt aatgctgatg ctaaaaaagt gactttgac 3240
 aagttaaag attcaaaaat ctcgactgac ggtcacaatg taacactaaa tagcgaagtg 3300
 aaaacgtcta atggtagtag caatgctggt aatgataaca gcaccgggtt aaccatttcc 3360
 gcaaaagatg taacggtaaa caataacgtt acctcccaca agacaataaa tatctctgccc 3420
 gcagcaggaa atgtaacaac caaagaaggc acaactatca atgcaaccac aggcagcgtg 3480
 gaagtaactg ctcaaaatgg tacaattaaa ggcaacatca cctcgcaaaa tgtaacagtg 3540
 acagcaacag aaaatctgt taccacagag aatgctgtca ttaatgcaac cagcggcaca 3600
 gtaaacatca gtacaaaaac agggatatt aaaggtggaa ttgaatcaac ttccggtaat 3660
 gtaaatatta cagcgagcgg caatacactt aaggttaagta atatcactgg tcaagatgt 3720
 acagtaacag cggatgcagg agccttgaca actacagcag gctcaaccat tagtgcgaca 3780
 acaggcaatg caaatattac aaccaaaaaca ggtgatatac acggtaaagt tgaatccagc 3840
 tccggctctg taacacttgt tgcaactggc gcaactctt ctgttaggtaa tatttcagg 3900
 aacactgtta ctattactgc ggatagcgtt aaattaacct ccacagtagg ttctacaatt 3960
 aatgggacta atagtgtaac cacctcaagc caatcagggc atattgaagg tacaatttct 4020
 ggttaatacag taaatgttac agcaagcact ggtgatttaa ctattggaaa tagtgcaaaa 4080
 gttgaagcga aaaatggagc tgcaacctta actgctgaat caggcaaaatt aaccacccaa 4140
 acaggctcta gcattacctc aagaatggt cagacaactc ttacagccaa ggatagcagt 4200
 atcgcaggaa acattaatgc tgctaatgtg acgtttaataa ccacaggcac tttactact 4260
 acaggggatt caaagattaa cgcaaccagt ggtaccttaa caatcaatgc aaaagatgcc 4320
 aaatttagatg gtgctgcac aggtgaccgc acagtagtaa atgcaactaa cgcaagtggc 4380
 tctggtaacg tgactgcgaa aacctaagc agcgtgaata tcaccgggg tttaaacaca 4440
 ataaatgggt taaatatcat ttcggaaaat ggttagaaaca ctgtgcgcctt aagaggcaag 4500
 gaaattgatg tgaaatatat ccaaccagggt gtagcaagcg tagaagaggt aattgaagcg 4560
 aaacgcgtcc ttgagaaggt aaaagattta tctgatgaag aaagagaaaac actagccaa 4620
 ctgggtgtaa gtgctgtacg tttcgtttag ccaaataatg ccattacggt taatacacaa 4680
 aacgagtttca acccaaaacc atcaagtcaa gtgacaattt ctgaaggtaa ggcgttcc 4740
 tcaagtggta atggcgcacg agtatgtacc aatgttgctg acgtatggaca gcag 4794

<210> 8

<211> 4803

<212> DNA

<213> Haemophilus influenzae

<400> 8

atgaacaaga tatatcgctc caaattcagc aaacgcctga atgctttggc tgctgtgtct 60
 gaattgacac ggggttgcgaa ccattccaca gaaaaaggca gtgaaaaacc tgttcgta 120
 aaagtacgccc acttggcggtt aaagccactt tccgctatata tgcatactttt gggcatggca 180
 tccattccgc aatctgtttt agcgagcgtt ttacaggggaa tgagcgtcgt acacggtaca 240
 gcaaccatgc aagtagacgg caataaaaacc actatccgtaa atagcgtcaat tgctatcatc 300
 aattggaaac aatttaacat tgacccaaat gaaatggtc agtttttaca agaaaggcagc 360
 aactctgcgcg tttcaaccgc tgttacatct gaccaaattt cccaaattttttaa agggattttt 420
 gattctaacg gacaagtctt tttaatcaac ccaaataggta tcacaatagg taaagacgca 480
 atttataaca ctaatggctt tactgctt acgcttagaca tttctaacga aacatcaag 540
 ggcgttaattt tcacccttgc gcaaccaaaag gataaaaggcac tcgctgaaat cgtaatcac 600
 ggtttaattt ccgttggtaa agacggtagc gtaaaaccttta ttggggcaaa agtggaaaaac 660
 gagggcgtga tttagcgtaa tggcggtgtt atttctttac ttgcaggggca aaaaatcacc 720
 atcagcgata taataaaatcc aaccatcaact tacagcattt ctgcacccgtaa aacgaagcg 780
 atcaatctgg gcgatatttt tgccaaagggt ggtacatcaat atgtccgcgc tgccactatt 840
 cgcacataaaag gtaaaacttgc tgccgactct gtaagcataag ataaaaatggg taacattgtt 900
 ctctctgcca aagaagggtga agcggaaattt ggcgggtgtaa tttccgctca aatcagca 960
 gccaaagggtg gtaagggtat gattacaggt gataaaaggta cattaaaaac aggtgcagg 1020
 atcgacccctt caggtaaaga agggggagag acttatcttgc ggggtatgaa ggcgtggcgaa 1080
 ggtaaaaatgt gtattcaattt agcgaagaaa acctcttttag aaaaaggctc gacaatttaat 1140
 gtatcaggca aagaaaaagg cggcgccgtt attgtatggg gcgatattgc attaatttaat 1200
 ggttaacatcaatcaagg tagcgatattt gctaaaaactg gcccgtttgtt gggaaacatca 1260
 ggacatgact tatccattgg tgatgtatgtt attgttgacg ctaaagagtg gttatttagac 1320

ccagatgatg tgcattgt aactttaca tctggacgca ataataccgg cgaaaaccaa 1380
 gatataca caggagatgg gactaaagag tcacctaag gtaatagtat ttctaaacct 1440
 acattaacaa actcaactct tgagcaaatc ctaagaagag gttcttatgt taatatcact 1500
 gctaataata gaatttatgt taatagctcc atcaacttat ctaatggcag ttaacactt 1560
 cacactaaac gagatggagt taaaattaac ggtgatatta cctcaaacga aatggtaat 1620
 ttaaccatta aagcaggctc ttgggttgc gttcataaaa acatcacgct tggtaacgggt 1680
 ttttgaata ttgcgtcgg ggattctgtc gctttgaga gagagggcga taaagcacgt 1740
 aacgaacacag atgctcaat taccgcacaa gggacgataa cctgtcaataa agatgataaa 1800
 caatttagat tcaataatgt atcttataac gggacgggca agggtttaaa gtttattgca 1860
 aatcaaaaata atttcactca taaatttgc ggcgaaatta acatatctgg aatagtaaca 1920
 attaaccaaa ccacgaaaaa agatgtaaa tactggaatg catcaaaaaga ctcttactgg 1980
 aatgtttctt ctcttacttt gaatacgggt caaaaattta ctttataaaa attcggttgc 2040
 agcggctcaa attcccaaga ttggaggtca tcacgtgaa gtttgcagg cgtacatttt 2100
 aacggcatcg gaggcaaaac aaacttcaac atcggagcta acgcaaaagc cttatttaaa 2160
 taaaacccaa acgcccgtac agacccaaaaa aaagaattac ctattacttt taacgccaac 2220
 attacagcta cccgttaacag tgatagctc gtgtatgtt acatacacgc caatcttacc 2280
 tctagagctg ccggcataaaa catggattca attaacattt cccggggct tgactttcc 2340
 ataacatccc ataatcgaa tagtaatgc ttggaaatca aaaaagactt aactataaat 2400
 gcaactggct cgaattttag tcttaagcaa acgaaagattt cttttataa tgaatacagc 2460
 aaacacgcca ttaactcaag tcataatcta accattctt cccggcaatgt cactcttaggt 2520
 gggggaaaattt caagcagtag cattacgggc aatataata tcaccaataa agcaaatgtt 2580
 acattacaag ctgacaccag caacagcaac acaggcttgc agaaaagaac tctaactctt 2640
 ggcataatata ctgttgaggg gaatttaagc ctaactgggt caaatgcaaa cattgtcgcc 2700
 aatctttcta ttgcagaaga ttccacattt aaaggagaag ccagtgcacaa cctaaacatc 2760
 accggcacct ttaccaacaa cggtaccggc aacattaata taaaacaagg agtggtaaaa 2820
 ctccaaggcg atattatcaa taaagggtgtt ttggaaatca ctactaacgc ctcaggcact 2880
 caaaaacca ttattaacgg aaatataact aacgaaaaag ggcacttaaa catcaagaat 2940
 attaaagccg acgcccgaat cccaaatttgc ggcataatct cacaaaaaaagg aggcaatctc 3000
 acaatttctt ctgataaagt aaatattacc aatcagataa caatcaaagc aggcgttgc 3060
 gggggcgtt ctgattcaag tgaggcagaa aatgctaacc taactattca aaccaaaagag 3120
 taaaatttgg caggagacct aaatatttca ggcttataa aagcagaaat tacagctaaa 3180
 aatggcagtg atttaactat tggcaatgcg acgggtggta atgctgatgc taaaaaaagt 3240
 actttgaca aggttaaaga ttcaaaaatc tcgactgcg gtcacaatgt aacactaaat 3300
 agcgaagtga aaacgtctaa tggtagtagc aatgctggta atgataacag caccggttt 3360
 accatttccg caaaagatgt aacggtaaac aataacgtt cctcccacaa gacaataaat 3420
 atctctgccc cagcagggaaa tggtaacaacc aaagaaggca caactatcaa tgcaaccaca 3480
 ggcagcgtgg aagtaactgc tcaaaaatgtt acaattaaaag gcaacattac ctcgcaaaaat 3540
 gtaacagtga cagcaacaga aaatcttgc accacagaga atgctgtcat taatgcaacc 3600
 agcggcacag taaacattag tacaaaaaaca ggggatatta aaggtggaaat tgaatcaact 3660
 tccggtaatg taaatattac agcagcggc aatacactt aagtaagttaa tatcactgg 3720
 caagatgtaa cagtaacagc ggtgcaggg gccttgacaa ctacagcagg ctcaaccatt 3780
 agtgcgacaa caggcaatgc aaatattaca accaaaacag gtgatattca cggtaaagg 3840
 gaatccagct cccgctctgt aacacttgc gcaactggag caactcttgc tggtaggtat 3900
 atttcaggta acactgttac tattactgcg gatagcggta aattaacactc cacagtaggt 3960
 tctacaatta atgggactaa tagtgcattt acctcaagcc aatcaggcga tattgaagg 4020
 acaatttctg gtaatacagt aaatgttaca gcaagcactg gtgatattac tattggaaat 4080
 agtgcggaaatg ttgaagcggaa aaatggagct gcaaccttta ctgctgaatc aggcaatatt 4140
 accacccaaa caggctctag cattacctca agcaatggc agacaactct tacagccaag 4200
 gatagcagta tcgcaggaaa cattatgc gctaattgtgc cgttaaatac cacaggcact 4260
 ttaactacta caggggattt aaagattaac gcaaccagggt gtacccttac aatcaatgca 4320
 aaagatgcca aatttagatgg tgctgcattt ggtgaccgc cagtagtaaa tgcaactaac 4380
 gcaagtggct ctggtaacgt gactgcggaa acctcaagca gcgtgaaat caccggggat 4440
 ttaacacaaa taaatgggtt aaatattttc tcggaaaaatg gtggaaacac tggtaacggtt 4500
 agaggcaagg aaattgtatgt gaaatataatc caaccagggt tagcaagcgt agaagaggt 4560
 attgcggcga aacgcgtcct tgagaaggtaa aaagattttt ctgatgaaaga aagagaaaca 4620
 ctggccaaac ttgggtgtaa tgctgtacgt ttgcgttgcg cttttatgc cttttacgggt 4680
 aatacacaacaa acgagtttac aaccacccaa tcaagtcaag tgacaatttc tgaaggtaag 4740
 gcgtgtttctt ctgggttgcgacca gttatgtacca atgttgcgtga cgatggacag 4800
 cag
 4803

<210> 9
 <211> 1599
 <212> PRT
 <213> Haemophilus influenzae

<400> 9
 Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Lys Asp Ile Ala Lys Thr Gly Gly Phe
 405 410 415
 Val Glu Thr Ser Gly His Tyr Leu Ser Ile Asp Asp Asn Ala Ile Val
 420 425 430
 Lys Thr Lys Glu Trp Leu Leu Asp Pro Glu Asn Val Thr Ile Glu Ala
 435 440 445
 Pro Ser Ala Ser Arg Val Glu Leu Gly Ala Asp Arg Asn Ser His Ser
 450 455 460
 Ala Glu Val Ile Lys Val Thr Leu Lys Lys Asn Asn Thr Ser Leu Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Ile Ser Asn Leu Leu Lys Ser Ala His Val
 485 490 495
 Val Asn Ile Thr Ala Arg Arg Lys Leu Thr Val Asn Ser Ser Ile Ser
 500 505 510
 Ile Glu Arg Gly Ser His Leu Ile Leu His Ser Glu Gly Gln Gly Gly
 515 520 525
 Gln Gly Val Gln Ile Asp Lys Asp Ile Thr Ser Glu Gly Gly Asn Leu
 530 535 540
 Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr Leu
 545 550 555 560
 Gly Ser Gly Phe Leu Asn Ile Thr Thr Lys Glu Gly Asp Ile Ala Phe
 565 570 575
 Glu Asp Lys Ser Gly Arg Asn Asn Leu Thr Ile Thr Ala Gln Gly Thr
 580 585 590

Ile Thr Ser Gly Asn Ser Asn Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Ser Leu Gly Gly Lys Leu Ser Phe Thr Asp Ser Arg Glu Asp Arg
 610 615 620
 Gly Arg Arg Thr Lys Gly Asn Ile Ser Asn Lys Phe Asp Gly Thr Leu
 625 630 635 640
 Asn Ile Ser Gly Thr Val Asp Ile Ser Met Lys Ala Pro Lys Val Ser
 645 650 655
 Trp Phe Tyr Arg Asp Lys Gly Arg Thr Tyr Trp Asn Val Thr Thr Leu
 660 665 670
 Asn Val Thr Ser Gly Ser Lys Phe Asn Leu Ser Ile Asp Ser Thr Gly
 675 680 685
 Ser Gly Ser Thr Gly Pro Ser Ile Arg Asn Ala Glu Leu Asn Gly Ile
 690 695 700
 Thr Phe Asn Lys Ala Thr Phe Asn Ile Ala Gln Gly Ser Thr Ala Asn
 705 710 715 720
 Phe Ser Ile Lys Ala Ser Ile Met Pro Phe Lys Ser Asn Ala Asn Tyr
 725 730 735
 Ala Leu Phe Asn Glu Asp Ile Ser Val Ser Gly Gly Ser Val Asn
 740 745 750
 Phe Lys Leu Asn Ala Ser Ser Asn Ile Gln Thr Pro Gly Val Ile
 755 760 765
 Ile Lys Ser Gln Asn Phe Asn Val Ser Gly Gly Ser Thr Leu Asn Leu
 770 775 780
 Lys Ala Glu Gly Ser Thr Glu Thr Ala Phe Ser Ile Glu Asn Asp Leu
 785 790 795 800
 Asn Leu Asn Ala Thr Gly Gly Asn Ile Thr Ile Arg Gln Val Glu Gly
 805 810 815
 Thr Asp Ser Arg Val Asn Lys Gly Val Ala Ala Lys Asn Ile Thr
 820 825 830
 Phe Lys Gly Gly Asn Ile Thr Phe Gly Ser Gln Lys Ala Thr Thr Glu
 835 840 845
 Ile Lys Gly Asn Val Thr Ile Asn Lys Asn Thr Asn Ala Thr Leu Arg
 850 855 860
 Gly Ala Asn Phe Ala Glu Asn Lys Ser Pro Leu Asn Ile Ala Gly Asn
 865 870 875 880
 Val Ile Asn Asn Gly Asn Leu Thr Thr Ala Gly Ser Ile Ile Asn Ile
 885 890 895

Ala Gly Asn Leu Thr Val Ser Lys Gly Ala Asn Leu Gln Ala Ile Thr
 900 905 910
 Asn Tyr Thr Phe Asn Val Ala Gly Ser Phe Asp Asn Asn Gly Ala Ser
 915 920 925
 Asn Ile Ser Ile Ala Arg Gly Gly Ala Lys Phe Lys Asp Ile Asn Asn
 930 935 940
 Thr Ser Ser Leu Asn Ile Thr Thr Asn Ser Asp Thr Thr Tyr Arg Thr
 945 950 955 960
 Ile Ile Lys Gly Asn Ile Ser Asn Lys Ser Gly Asp Leu Asn Ile Ile
 965 970 975
 Asp Lys Lys Ser Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln
 980 985 990
 Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr Asn
 995 1000 1005
 Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser Ser
 1010 1015 1020
 Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys Leu
 1025 1030 1035 1040
 Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala
 1045 1050 1055
 Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn Ala
 1060 1065 1070
 Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser Asn
 1090 1095 1100
 Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile Ser
 1105 1110 1115 1120
 Ala Lys Asp Val Thr Val Asn Asn Val Thr Ser His Lys Thr Ile
 1125 1130 1135
 Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr Thr
 1140 1145 1150
 Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly Thr
 1155 1160 1165
 Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr Glu
 1170 1175 1180
 Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly Thr
 1185 1190 1195 1200
 Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu Ser
 1205 1210 1215

Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys Val
 1220 1225 1230
 Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly Ala
 1235 1240 1245
 Leu Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn Ala
 1250 1255 1260
 Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser Ser
 1265 1270 1275 1280
 Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val Gly
 1285 1290 1295
 Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys Leu
 1300 1305 1310
 Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr Thr
 1315 1320 1325
 Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr Val
 1330 1335 1340
 Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala Lys
 1345 1350 1355 1360
 Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly Lys
 1365 1370 1375
 Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln Thr
 1380 1385 1390
 Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala Ala
 1395 1400 1405
 Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp Ser
 1410 1415 1420
 Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp Ala
 1425 1430 1435 1440
 Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala Thr
 1445 1450 1455
 Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Val
 1460 1465 1470
 Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile Ser
 1475 1480 1485
 Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp Val
 1490 1495 1500
 Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu Ala
 1505 1510 1515 1520

Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg Glu
 1525 1530 1535

Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro Asn
 1540 1545 1550

Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro Ser
 1555 1560 1565

Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly Asn
 1570 1575 1580

Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595

<210> 10

<211> 1600

<212> PRT

<213> Haemophilus influenzae

<400> 10

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Asp Ile Ala Lys Thr Gly Gly Phe
 405 410 415
 Val Glu Thr Ser Gly His Asp Leu Ser Ile Gly Asp Asp Val Ile Val
 420 425 430
 Asp Ala Lys Glu Trp Leu Leu Asp Pro Asp Asp Val Ser Ile Glu Thr
 435 440 445
 Leu Thr Ser Gly Arg Asn Asn Thr Gly Glu Asn Gln Gly Tyr Thr Thr
 450 455 460
 Gly Asp Gly Thr Lys Glu Ser Pro Lys Gly Asn Ser Ile Ser Lys Pro
 465 470 475 480
 Thr Leu Thr Asn Ser Thr Leu Glu Gln Ile Leu Arg Arg Gly Ser Tyr
 485 490 495
 Val Asn Ile Thr Ala Asn Asn Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510

Leu Ser Asn Gly Ser Leu Thr Leu His Thr Lys Arg Asp Gly Val Lys
 515 520 525
 Ile Asn Gly Asp Ile Thr Ser Asn Glu Asn Gly Asn Leu Thr Ile Lys
 530 535 540
 Ala Gly Ser Trp Val Asp Val His Lys Asn Ile Thr Leu Gly Thr Gly
 545 550 555 560
 Phe Leu Asn Ile Val Ala Gly Asp Ser Val Ala Phe Glu Arg Glu Gly
 565 570 575
 Asp Lys Ala Arg Asn Ala Thr Asp Ala Gln Ile Thr Ala Gln Gly Thr
 580 585 590
 Ile Thr Val Asn Lys Asp Asp Lys Gln Phe Arg Phe Asn Asn Val Ser
 595 600 605
 Leu Asn Gly Thr Gly Lys Gly Leu Lys Phe Ile Ala Asn Gln Asn Asn
 610 615 620
 Phe Thr His Lys Phe Asp Gly Glu Ile Asn Ile Ser Gly Ile Val Thr
 625 630 635 640
 Ile Asn Gln Thr Thr Lys Lys Asp Val Lys Tyr Trp Asn Ala Ser Lys
 645 650 655
 Asp Ser Tyr Trp Asn Val Ser Ser Leu Thr Leu Asn Thr Val Gln Lys
 660 665 670
 Phe Thr Phe Ile Lys Phe Val Asp Ser Gly Ser Asn Gly Gln Asp Leu
 675 680 685
 Arg Ser Ser Arg Arg Ser Phe Ala Gly Val His Phe Asn Gly Ile Gly
 690 695 700
 Gly Lys Thr Asn Phe Asn Ile Gly Ala Asn Ala Lys Ala Leu Phe Lys
 705 710 715 720
 Leu Lys Pro Asn Ala Ala Thr Asp Pro Lys Lys Glu Leu Pro Ile Thr
 725 730 735
 Phe Asn Ala Asn Ile Thr Ala Thr Gly Asn Ser Asp Ser Ser Val Met
 740 745 750
 Phe Asp Ile His Ala Asn Leu Thr Ser Arg Ala Ala Gly Ile Asn Met
 755 760 765
 Asp Ser Ile Asn Ile Thr Gly Gly Leu Asp Phe Ser Ile Thr Ser His
 770 775 780
 Asn Arg Asn Ser Asn Ala Phe Glu Ile Lys Lys Asp Leu Thr Ile Asn
 785 790 795 800
 Ala Thr Gly Ser Asn Phe Ser Leu Lys Gln Thr Lys Asp Ser Phe Tyr
 805 810 815

Asn Glu Tyr Ser Lys His Ala Ile Asn Ser Ser His Asn Leu Thr Ile
 820 825 830
 Leu Gly Gly Asn Val Thr Leu Gly Gly Glu Asn Ser Ser Ser Ile
 835 840 845
 Thr Gly Asn Ile Asn Ile Thr Asn Lys Ala Asn Val Thr Leu Gln Ala
 850 855 860
 Asp Thr Ser Asn Ser Asn Thr Gly Leu Lys Lys Arg Thr Leu Thr Leu
 865 870 875 880
 Gly Asn Ile Ser Val Glu Gly Asn Leu Ser Leu Thr Gly Ala Asn Ala
 885 890 895
 Asn Ile Val Gly Asn Leu Ser Ile Ala Glu Asp Ser Thr Phe Lys Gly
 900 905 910
 Glu Ala Ser Asp Asn Leu Asn Ile Thr Gly Thr Phe Thr Asn Asn Gly
 915 920 925
 Thr Ala Asn Ile Asn Ile Lys Gly Val Val Lys Leu Gly Asp Ile Asn
 930 935 940
 Asn Lys Gly Gly Leu Asn Ile Thr Thr Asn Ala Ser Gly Thr Gln Lys
 945 950 955 960
 Thr Ile Ile Asn Gly Asn Ile Thr Asn Glu Lys Gly Asp Leu Asn Ile
 965 970 975
 Lys Asn Ile Lys Ala Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr
 995 1000 1005
 Asn Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser
 1010 1015 1020
 Ser Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn
 1060 1065 1070
 Ala Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile
 1075 1080 1085
 Ser Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser
 1090 1095 1100
 Asn Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile
 1105 1110 1115 1120
 Ser Ala Lys Asp Val Thr Val Asn Asn Val Thr Ser His Lys Thr
 1125 1130 1135

Ile Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr
 1140 1145 1150

 Thr Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly
 1155 1160 1165

 Thr Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr
 1170 1175 1180

 Glu Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly
 1185 1190 1195 1200

 Thr Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu
 1205 1210 1215

 Ser Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys
 1220 1225 1230

 Val Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly
 1235 1240 1245

 Ala Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn
 1250 1255 1260

 Ala Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser
 1265 1270 1275 1280

 Ser Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val
 1285 1290 1295

 Gly Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys
 1300 1305 1310

 Leu Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr
 1315 1320 1325

 Thr Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr
 1330 1335 1340

 Val Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala
 1345 1350 1355 1360

 Lys Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly
 1365 1370 1375

 Lys Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln
 1380 1385 1390

 Thr Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala
 1395 1400 1405

 Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp
 1410 1415 1420

 Ser Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp
 1425 1430 1435 1440

Ala Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala
1445 1450 1455

Thr Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser
1460 1465 1470

Val Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile
1475 1480 1485

Ser Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp
1490 1495 1500

Val Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu
1505 1510 1515 1520

Ala Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg
1525 1530 1535

Glu Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro
1540 1545 1550

Asn Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro
1555 1560 1565

Ser Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly
1570 1575 1580

Asn Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
1585 1590 1595 1600

<210> 11

<211> 29

<212> PRT

<213> Haemophilus influenzae

<400> 11

Val Asp Glu Val Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp
1 5 10 15

Leu Ser Asp Glu Glu Arg Glu Ala Leu Ala Lys Leu Gly
20 25